THE EFFECTS OF VARYING THE NUMBER OF CUES IN A MULTIPLE-CUE INFERENCE TASK

A THESIS

Presented to

The Faculty of the Division of Graduate

Studies and Research

By

Alan Leslie Dorris

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

in the School of Industrial and Systems Engineering

Georgia Institute of Technology

December, 1974

THE EFFECTS OF VARYING THE NUMBER OF CUES

IN A MULTIPLE-CUE INFERENCE TASK

Approved:

T. Connolly, Co-Chairman

T. L. Sadosky, Co-Chairman

E. J. Baker

D. C. Montgomery

A. O. Esogbue

Date approved by Chairman: Nov. 1, 1974

ACKNOWLEDGMENTS

Each member of my dissertation reading committee made substantive contributions to my research. Dr. E. J. Baker introduced me to the area of human information processing and provided suggestions regarding the research. The data analysis and experimental design phases were contributed to by Dr. D. C. Montgomery, who pointed out several deficiencies in the original draft of this dissertation. Dr. A. O. Esogbue shared his enthusiasm for both prescriptive and descriptive decision models and made helpful suggestions which improved the research.

Dr. Terry Connolly and Dr. Thomas Sadosky served as co-chairmen for the committee and provided excellent guidance. Dr. Connolly shared his enthusiasm for the study of human decision processes and introduced me to the lens model paradigm which was employed in this research. His intuitive understanding and technical knowledge were of immense value throughout the course of the research. Dr. Sadosky was my academic advisor during graduate school and guided my work in the area of human factors. His interest in the study of human behavior in general and in my work in particular were of value. Both of these men were friends as well as teachers and were always available for discussion or advice. It would be difficult to overestimate their contributions to my graduate studies. I welcome this opportunity to express my sincere gratitude to them.

It is customary to acknowledge at this point the contributions of one's wife or husband. In my case, I can only hope that my wife, Patsy, realizes that I appreciate the sacrifices that she made during the years spent in school. I could never have attended graduate school without her support, and I sincerely hope that she will have an opportunity to do some of the things that she has had to postpone on my behalf. This dissertation is dedicated to her and our son, Jason.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ILLUSTRATIONS

SUMMARY

The general area addressed by this research is that of determining the effects of the amount of information provided to a human decision maker on the quality of the decisions made. A distinction is drawn between providing more information to the decision maker and thereby facilitating the process and providing more data which may tend to cause distractions and thus impede the process. It is argued that an understanding of the effects of information or data overload situations may lead to improved designs for decision making systems such as management information systems. In addition, it will lead to a fuller understanding of the human information processing system.

The research reported here consists of a series of studies conducted in a laboratory setting within the framework of the lens model paradigm. The lens model is a conceptualization of the human inference process which has been found useful in the study of human judgment. The model is described in detail in the report.

The studies which are presented here examine the baseline condition of performance when several pieces of information or cues are available and the cues are of equal validity. Performance is observed as the number of cues is varied under differing sets of conditions. Among the task variables considered are the degree of cue intercorrelation or redundancy, cue validity, task predictability and amount of structuring insight provided. The experiments are the first known studies involving varying numbers of cues in a context free multiple cue inference task.

The results of these studies indicate that although people are able to utilize increased amounts of information, increased amounts of data may affect decision making behavior in non-intuitive fashions. In particular, increased amounts of data which are not accompanied by increased task predictability may result in a higher level of performance to a certain point with a subsequent decrement in performance as more data are presented. Suggestions for further research in this area are made and the methodological problems inherent in studies involving varying amounts of information are discussed.

CHAPTER I

INTRODUCTION TO JUDGMENT LITERATURE AND RESEARCH PROBLEM DEFINITION

Characteristics of Judgment Problems

According to Ellis [1972], "Judgment, as a perceptual response indicator, refers simply to the placing or ordering of stimuli along some scale." Thus, the child learns that two quite different objects are both "chairs" and the taxonomist classifies apes and horses as "mammals." These are examples of concept formation (see Bruner, Goodnow and Austin [1956]) and involve the use of a nominal scale. The smoker may prefer pipes to cigars and thus employ a preference scale which may have only ordinal properties. Similarly, a stock market analyst predicts that stock A will advance ten points this week and uses an interval (or ratio) scale.

Clearly a rather diverse class of human activities employs judgment as defined above. Indeed, a great number of occupations, such as that of the pathologist or of the television weather-man, are concerned primarily with the exercising of judgment. Even with the acceptance of the definition quoted above, however, it is not obvious what the components of judgment are. It is of some importance and interest, then, that a brief exploration of the process be presented.

In an informative discussion Newell [1968] characterizes judgment as a cognitive process involving an available set of inputs and a well defined domain of outputs. The process, according to Newell, is not extended in time and is neither a simple transduction of information nor the application of a given rule. By these standards, then, information search and creative thinking are not judgment. Thus, it may be convenient to distinguish the term "judgment" from the more general term "decision-making" which may involve the application of a given rule. However, unless it is explicitly stated otherwise, "judgment" and "decision-making" will be used interchangeably in this research.

A somewhat more general characterization of the judgment process is presented by Einhorn [1974] who proposes three components of the judgmental process. The first stage refers to the identification and organization of relevant information in the environment. This, of course, implies at least a limited amount of search behavior. The measurement or evaluation of the relevant information or cues composes the second stage. Finally, cues must be weighted and combined in some fashion.

The process is illustrated well by a hypothetical example from medical diagnosis. In performing an examination, the physician will have access to certain standard types of information, such as blood pressure, obesity or shortness of breath. Additionally, some less obvious cues may be utilized, such as speech patterns or manner of walking. Having identified and clustered the cues, they are measured or evaluated. Some, such as blood count and temperature, are subject to objective measurement, whereas others, for example nervous tremor, are evaluated subjectively. In any case, however, the process defined above by Ellis,

". . . the placing or ordering of stimuli along some scale," is involved. Finally, an overall or global judgment is required, such as whether or not a malignancy is present or whether surgery is required.

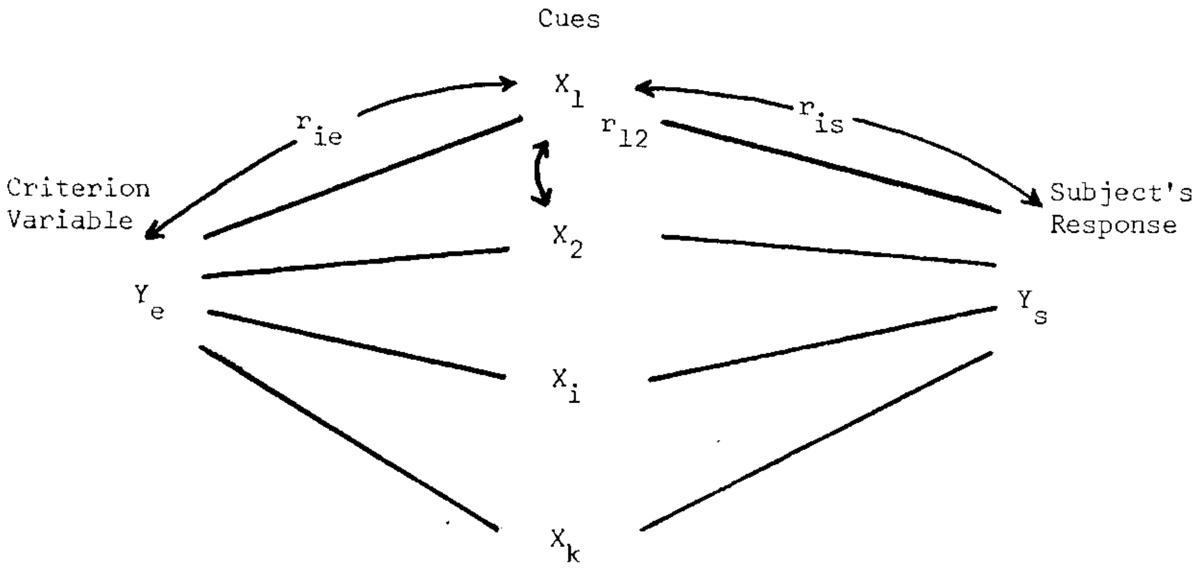
A consideration of typical judgment tasks yields several characteristics common to such problems. Perhaps the most notable aspect is that none of the information available to the decision maker is sufficient to make exact predictions on each problem. Another way of stating this is that each piece of information is only probabilistically related to the variable to be predicted. It is also true that all of the information taken together is not sufficient to predict with perfect accuracy. Thus, there is an irreducible random error component which means that while decision makers may have varying degrees of success, none can be perfectly accurate. There may well be cases where this is not the case, but those cases are of little interest to the study of decision-making performance. The main point is that any model of the judgment process must include these elements at a minimum.

A simplified conceptual model of the judgment process would involve some underlying criterion variable about which inference is to be drawn, a set of cues relevant to the criterion variable and the judge's response to the cues or his estimate of the criterion variable. The criterion variable may be thought of as a distal variable while the cues are proximal stimuli. It seems clear that this model is a simplification of the general model which would include many difficulties, such as multi-dimensional criteria, search for new cues on the basis of previously perceived ones, and the formulation of sub-hypotheses on the basis of subsets of the cues. It is also true that the perception and evaluation of individual or subsets of cues may be considered as judgment processes in themselves. The simplified model described here, however, is of interest and of sufficient generality to guide research efforts.

The Lens Model

The conceptual model discussed above has been formalized by Brunswik [1955]. The fundamental concept of this representation of the judgment problem is that the overall system must consist of a task sub-system and a response sub-system. This crucial point allows performance to be evaluated with respect to the particular task being undertaken. Thus, performance is evaluated on both an absolute basis (comparison of judgment with true value of criterion variable) and on a relative basis (comparison of judgment with the best estimate of the criterion variable). An additional feature is that mathematical-statistical models can be constructed of both the environment and the decision-maker thus allowing for comparisons of the environment, the model of the environment, the judge and the model of the judge.

Brunswik's model may be schematically represented by the drawing in Figure 1. The environmental criterion variable, Y_e, is to be judged, estimated or predicted on the basis of several cues, each of which is presumed to furnish some information about Y_e. Additionally, the cues are, in general, partially redundant and thus a given cue, X_i, cannot be regarded as giving information about Y_e which is independent of information provided by some other cue, X_j. It is clear that in some



$$\hat{Y}_e = \beta_0 + \beta_1 X_1 + \ldots + \beta_k X_k \qquad \hat{Y}_s = \beta_0^* + \beta_1^* X_1 + \ldots + \beta_k^* X_k$$

Figure 1. The Lens Model

circumstances two pieces of inforation, X_i and X_j, both relevant to Y_e, will be statistically independent; however, this is not true in general.

The task presented to the subject, then, is one of perceiving, measuring and combining the cues so as to produce an overall judgment, Y_s, of Y_e. This process of relevant information emanating from some criterion variable and being recombined into a judgment is analogous to light being focused by a convex lens. This analogy has led to the conceptual model under discussion being called the lens model.

Several options are available with regard to quantifying the relationships between criterion, cues and judgment and with respect to defining operationally the concepts of relevance and redundancy. One approach is to employ an information theoretic model. Another option is to use correlations between the several variables and define relevance in terms of cue validities and redundancy in terms of cue intercorrelations. It is this latter approach which has been traditionally employed and will be utilized in this research. It is thus assumed that the independent variables are random variables.

Returning to Figure 1, let $r_{ie} = r_{X_i,Y_e}$ be the validity of the *ith* cue, $r_{is} = r_{X_i,Y_s}$ be the utilization coefficient of the *ith* cue and $r_{ij} = r_{X_i,X_j}$ be the correlation between cues X_i and X_j. For the special case of uncorrelated cues, i.e., $r_{ij} = 0.0$ for all $i \neq j$, r_{ie}^2 is a measure of the proportion of variance in Y_e which is explained by X_i and, similarly, r_{is}^2 is the proportion of variance in Y_s explained by X_i. For the more general case of correlated cues, the relationships are more complex.

A simple approach to predicting values of Y_e for observed values of the X_i would be to formulate a first-order regression model of Y_e on the X_i. The form of such a model would be:

$$\hat{Y}_e = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_K X_K. \tag{1}$$

The multiple correlation coefficient corresponding to Equation (1) is $R_e = r_{Y_e,\hat{Y}_e}$ and R_e^2 is a measure of the proportion of variance in Y_e which is explained by Equation (1). Similarly, a model of the subject's responses can be constructed by a first-order regression model of Y_s on the X_i resulting in:

$$\hat{Y}_s = \beta_0^* + \beta_1^* X_1 + \beta_2^* X_2 + \ldots + \beta_K^* X_K. \tag{2}$$

The β_i^*'s are not identical with the β_i's above. Again, $R_s = r_{Y_s,\hat{Y}_s}$ is the multiple correlation coefficient and R_s^2 is the proportion of variance in the responses which is explained by Equation (2).

Writing Y_e and Y_s in the following forms:

$$Y_e = \hat{Y}_e + Z_e \tag{3}$$

$$Y_s = \hat{Y}_s + Z_s \tag{4}$$

it is apparent that Y_e consists of a component which is explained by a first-order model and a residual component. A similar statement holds for Y_s. The form of Z_e and Z_s is not known in general. The residual may contain only random error in which case the first-order model accounts for all the explainable variation. On the other hand, the residual may contain "lack-of-fit" variation which reflects the fact that there is systematic variation due to terms not included in the first-order model. The extent to which Equations (1) and (2) are adequate descriptors of the environmental and response subsystems, respectively, is an empirical questions and subsequent discussion will indicate that first-order models perform admirably.

It is not surprising that the various correlation and regression coefficients associated with the lens model are functionally related. It would be desirable to express some performance measure as a function of various coefficients and parameter values from the model. Early work in this regard was conducted by Hursch, Hammond and Hursch [1964], Hammond, Hursch and Todd [1964] and subsequently by Tucker [1964]. Tucker's version, which has recently been examined in detail by Castellan [1973], will be employed here. In this form:

$$r_a = GR_eR_s + C(1-R_e^2)^{1/2}(1-R_s^2)^{1/2} \tag{5}$$

where:

$$r_a = r_{Y_e,Y_s} \tag{6}$$

$$G = r_{\hat{Y}_e,\hat{Y}_s} \tag{7}$$

$$C = r_{Z_e,Z_s}. \tag{8}$$

The *achievement* index, r_a, is the correlation between the value of the criterion variable, Y_e, and the judge's response, Y_s. As discussed above, R_e is a measure of the extent to which the first-order model of the environment, Equation (1), fits the values of Y_e. R_e is referred to as the *task predictability.* Similarly, R_s is a measure of the fit of Equation (2) to the responses and is termed *response linearity.*

G, as defined in Equation (7), is a measure of the extent to which Equation (1) matches Equation (2) in the sense that the regression coefficients of one model are proportional to those of the other model. It is clear that if each regression coefficient of Equation (1) can be written as the same positive linear transformation of the corresponding regression coefficient in Equation (2), then G = 1.0. It is very important to note that a G value of unity implies proportional weights in the environmental and response systems only to the extent that first-order models apply. Thus, a first-order model of the response system may match that of the environment well, i.e., G close to 1.0, and yet the first-order model be a very poor description of the responses, i.e., R_s close to 0.0. Reasoning similar to this has led Hammond and Summers [1972] to refer to G as *knowledge* and R_s as *cognitive control* implying that the subject may correctly detect the first-order weighting policy but fail to utilize this policy consistently.

The remaining quantity in Equation (5) is C, the correlation between the residual in the environment and the residual in the responses. Clearly, if either Z_e or Z_s is completely random, then C will equal zero.

Thus, if all predictable variation in Y_e is explained by Equation (1), there is no component of Z_e which can be systematically utilized by the judge and C = 0.0. Similarly, if the subject's responses are described by Equation (2) plus a purely random component, then even if there is systematic variation in Z_e, e.g., terms of higher order, the judge is not using it and again C = 0.0. It follows that C will depart from zero only if there is systematic variation in Z_e which is employed by the judge.

The lens model Equation (5) relates achievement to the remaining quantities discussed above. As was previously mentioned, this version of the equation was originally presented by Tucker [1964] and elaborated upon by Castellan [1973]. Appendix A contains a derivation of Equation (5).

A special case of particular interest occurs when the environment is of first-order or when it is realistic to assume that the first-order model of the judge provides an adequate fit. In either case, the value of C will approach zero and the lens model Equation (5) reduces to:

$$r_a \cong G\, R_e R_s. \tag{9}$$

Of the quantities in Equation (9), R_e is not under the control of the judge since it is fixed by the statistical properties of the task sub-system. On the other hand, both G and R_s reflect the responses of the judge. In order to maximize r_a, both G and R_s should approach their maximum values of unity. This, of course, implies that the judge has correctly detected the first-order task properties and used them in a perfectly consistent manner. It follows immediately that $r_a \le R_e$.

Comments on the Lens Model

There is reason to question the soundness of the correlation-regression approach of the lens model. At best the measures are rather crude and provide only a macroscopic view of performance. G and R_s are particularly suspect and their interpretation is not nearly as straightforward as their derivation. R_s is a measure of the fit of the model to the subject's responses. This, however, assumes that the same model is appropriate throughout the block of trials. Thus, a subject who shifts from one policy or weighting scheme to another may not be fit well by a single model even if he employs the policies consistently while they are in use. This, of course, is a particular problem when subjects are learning a new task and receiving feedback about the results of previous judgments.

Dawes and Corrigan [1974] have suggested that most judgment tasks are of such a nature that any first order model will correlate highly with the optimal one. This means that G will tend to be large even when the judge uses a policy quite different from the best policy. Thus, high values of G must not be taken as indicative that the subject is necessarily using a model very similar to the best one.

Perhaps the most serious trap against which the analyst must guard is that of speaking of the regression model of the subject's responses as if it were a model of the subject's cognitive processes. It is not. The regression model is no more than a simulation of the overt response system of the judge with a structure which does not necessarily represent any conscious or unconscious mental processes. Although this point has been made previously, it is well worth emphasizing since the analyst should never lose sight of what is being modeled.

To this point, the entire discussion has assumed a first-order representation of both the environment and the judge. There is, of course, no reason why the above discussion could not be extended to models of higher order. A considerable amount of empirical evidence suggests that even though the judge, for example, may be using data in a highly non-linear fashion, the first-order model provides a satisfactory representation. This empirical finding has greatly influenced the methodology and direction of judgment research.

Purpose of the Research

Judgmental decisions are made under a variety of conditions. There may be a lack of relevant information or an overabundance; there may be time pressures or monetary constraints; the costs for failure may be high and the rewards for success great. In many cases, only qualitative information is available when quantitative information would be desirable. Feedback regarding the results of previous decisions may be delayed or wrong or nonexistent.

In order to understand how people make decisions in situations such as these, it is necessary to examine the effects of each of the difficulties cited above and the many others which make real-world decision making painful. It is not sufficient to observe behavior under a given set of conditions; relevant factors must be varied and the corresponding changes--or lack of change--noted. This is the case whether one has a theoretical or an applied interest.

The particular aspect of the inference process to be examined in this research is the effect of varying the number of cues available to the decision maker. The primary interest is in the effects of increased amounts of information on decision-making performance. The work is motivated by both theoretical and practical aspects. For example, a thorough understanding of the phenomena associated with information overload situations would be useful in the design of complex man-machine decision making systems such as management information systems. Such knowledge could result in a more effective allocation of information processing tasks between man and computer and, perhaps, even to the design of systems tailored to the specific abilities of the particular individual.

On the other hand, there are scientific or theoretical reasons for being concerned with the relationship between information input and decision-making performance. The classic paper by Miller [1956] has shown that man has a limited capacity for processing certain forms of information. Thus, there is a point beyond which additional information input will lead to no further increase in performance. It is of interest to explore the relationship of findings such as this to performance in the multiple cue inference tasks used in this research. One wishes to discover how the human subject attempts to cope with such conditions and what strategies are adopted to decrease the pressures of coping with large amounts of information.

Of course, the scope of any research undertaking is, of necessity, limited by time and resources. In the present case, interest is restricted to a set of multiple cue inference tasks where the number of cues available to the decision maker is varied. All studies are conducted in a laboratory setting so that the results are the product of carefully controlled conditions.

In summary, this research is conducted within the conceptual framework of the lens model paradigm with the intent of discovering how performance in a multiple cue inference task is affected by the number of cues. The results provide greater insight into man's limited information processing capacity, suggest implications for the design of decision making systems and point out areas of interest for further research.

Review of the Literature

In this section, a review of the relevant judgment literature will be presented. The initial part of the review will present a rather broad over-view of judgment research and examine the major findings. The latter phase of the review will focus specifically on the problem considered in this research--that of changes in performance as a function of varying the number of cues.

It is important to note the restrictions of the scope of this review. In the first place, it is not intended to be a comprehensive review but rather to introduce the area and concentrate on the problem of interest in this research. The interested reader is referred to Slovic and Lichtenstein [1971] for a more complete overview of this field. Secondly, the review presented here will concentrate almost exclusively on studies either conducted within the framework of the lens model or employing a similar methodology. This restriction, for example, will exclude Bayesian studies and information theoretic approaches. This does not imply that these avenues of research are unimportant or irrelevant to the lens model paradigm. The problem is rather that studies conducted within a different conceptual and methodological framework generally focus on different tasks and the results are not readily interpretable outside of the particular vein in which they were conducted. Again, the interested reader is referred to Slovic and Lichtenstein [1971] for a comparative review of several research paradigms.

For purposes of this review, it is convenient to divide the judgmental process into two phases--perception and measurement of cues and weighting and combining of cues. In most real-world judgment tasks, it is not possible to isolate all of the cues that might be employed by the decision maker. For example, in job interview situations, the interviewer may consciously or unconsciously utilize his perception of the job applicant's accent or his personal appearance or mannerisms. In the setting of a clinical diagnostic interview, Borke and Fiske [1957] controlled the availability of cues by allowing judges to (a) conduct a direct interview, (b) see and hear an interview through a one-way screen, (c) listen to a tape recording of an interview, and (d) read a verbatim transcript. The findings revealed no significant differences between the conditions which suggests that the additional information was offset by additional distractions. Luft [1951] found that for a projective test, groups of psychiatrists, social workers and clinical psychologists who read transcripts were outperformed by undergraduate students who listened to tape recordings.

Meehl [1954] sparked considerable research into the problem of comparing predictions made by expert judges with those made by purely statistical techniques such as multiple regression. The general finding, about which more will be said later, is that statistical methods are superior to clinical judgment. It has been pointed out, however, by Sawyer [1966] and Einhorn [1972] that the acts of deciding which variables to enter into the statistical model and of measuring the levels of those variables for a particular problem may be viewed as components of judgment. In particular, this is the initial phase--perception and measurement of cues--as pointed out above. Thus an overall or global judgment can be considered to be the result of a series of smaller component processes.

Einhorn [1972] reports a study in which pathologists were asked to predict the survival time (in months) for patients suffering from Hodgkin's disease on the basis of the relative amounts of nine histological signs from biopsy slides. The physicians rated each sign separately and also gave a global judgment of disease severity. The resulting analysis was inconclusive in that for some judges, addition of the global judgment hurt prediction whereas in other cases it helped. The implication is that although mechanical, i.e., statistical, combination of cues may be superior, the human may be quite effective at identifying relevant predictor information. This action may be viewed as a type of judgmental decision or, in a larger context, as a part of the overall judgment process.

The book by Meehl [1954] mentioned above led to considerable investigation of the relative validities of clinical judgment and of statistical predictions. Studies of this type are best illustrated by an example. For this purpose, the results of Goldberg [1965] will be examined.

The task involves classification of patients as neurotic or psychotic on the basis of Minnesota Multiphasic Personality Inventory (MMPI) profiles. Each profile consists of 11 scores on different scales and the interpretation of profiles has been considered to be a complex task involving configural information usage. The 861 patients whose profiles were employed had been diagnosed as either neurotic or psychotic on the basis of more extensive examination. Twenty-nine trained clinical psychologists diagnosed each of the 861 patients. In addition, a multiple regression (first-order) analysis was conducted to make diagnoses. The average correlation between clinical judgments and the binary criterion variable was 0.28 whereas the cross-validated validity of the regression model was 0.46. This finding is quite well established for a rather wide variety of tasks (see Dawes [1971, 1972] and Wiggins and Kohen [1971]).

It should be noted that the above problem can be expressed in terms of the quantities involved in the lens model equation. In particular, if R_e, the correlation between Y_e and $\hat{Y}_e$, exceeds r_a, the correlation between Y_e and Y_s, the model of the environment has outperformed the man; if r_a exceeds R_e, the man has outperformed the model of the environment. Clearly, since the $\hat{Y}_e$ values are based on the optimal (least-squares) model among the class of first-order models, the judge can outperform the model only if he can correctly and consistently utilize information in a non-linear, configural manner.

The discussion to this point has centered on a comparison of the judge's responses with the criterion variable and the model of the task subsystem. The lens model, however, is symmetric and the predictive power of the first-order representation of the response subsystem remains to be demonstrated. Further, it is necessary to identify particular factors which influence inference behavior.

Hoffman [1960] proposed that a regression of the judge's responses on the cues could be used to describe the judge's policy. He terms this model a "paramorphic representation" implying that it did not model the actual cognitive processes but rather it described the outcomes. It may be thought of as a simulation model whose structure is of little consequence but whose outputs resemble those of the activity being modeled.

The central finding of "policy-capturing" research is that first-order models of the form of Equation (2) do a very adequate job of modeling the judge. The general approach has been to fit a first-order model, compute the multiple correlation coefficient and add higher order terms of the form X_i^2 or X_iX_j only if they significantly increase the multiple correlation coefficient. An excellent example of this approach is provided by Wiggins and Hoffman [1968] who re-examined the MMPI data discussed above. The most compelling indication of the power of the first-order model is that, for the most configural judge, the increase in correlation between the model and his judgment, obtained by adding all terms of the form X_i^2 and X_iX_j, was 0.03. This seems to be a quite general finding for the case where each independent variable is conditionally monotonically related to the dependent variable. Dawes [1972] and Goldberg [1968] discuss a variety of studies which have confirmed this finding.

Having established that first-order models fit judges quite well for a wide variety of tasks, two questions must be answered: "Why do first-order models fit well even for tasks that subjects claim to be performing with complex information usage?" and "Does this mean that people do not use information configurally?" In reply to the first question, Yntema and Torgerson [1961] suggested more than a decade ago that for situations where the independent variables bear a conditionally monotonic relationship with the dependent variable, a first-order model will always fit well. They back this statement up with an example of a highly configural task in which 94 per cent of the variance was accounted for by a first-order model. Recently, Dawes and Corrigan [1974] have presented an expanded discussion of the prevalence of first-order models. A complete answer to the first question is not yet available.

The second question has received a considerable amount of attention. In the study by Wiggins and Hoffman [1968] cited above, they concluded that 16 of the 29 judges could be classified as "configural" even though configural models did little better than first-order models. Studies by Hoffman, Slovic and Rorer [1968] and Slovic [1969] have employed an analysis of variance approach to demonstrate non-linear usage of data.

The majority of the evidence indicates that the findings of Yntema and Torgerson are correct. The first-order model is an extremely powerful tool for description even when higher order terms are present in the actual cognitive processes employed. This finding has been a mixed blessing for researchers in that by describing complex processes in simple terms, the linear model suppresses complex information usage patterns which may be present.

Recent extensions of the above work (see Goldberg [1970], Dawes [1972] and Wiggins and Kohen [1971]) have concluded that in some cases the model of the judge may be more valid than the judge himself. This phenomenon has been called "bootstrapping" and has been uncovered in several judgment tasks. Dawes [1972] has found evidence that models with unit weights and even with randomly chosen weights may be more valid than the judge himself.

The research examined thus far has been on a rather macroscopic level and has focused on describing judgmental policies and comparing the judge with various models. Many studies have taken a more macroscopic approach and attempted to analyze the factors which may affect performance in a judgment task. For example, Slovic [1966] has observed that when two cues agree in their implications, judges tend to use them both. When, however, they are inconsistent, the judge ignores one and utilizes the other or he utilizes other cues.

Naylor and Schenck [1968] have examined the effects of cue redundancy, r_{12}, in a two-cue problem. They found that achievement is affected by task predictability, R_e, and by cue redundancy; further, there is an interaction effect between R_e and r_{12}. In particular, task predictability is more important under conditions of high cue redundancy than under conditions of low cue redundancy.

Dudycha, Dumoff and Dudycha [1973] have examined the effects of shifting the ecological validity in a one cue experiment. They examined a shift from high to low ecological validity and a shift from low to high validity. They found that prior exposure to a low validity task depressed both achievement and consistency in a subsequent high validity task from the levels obtained when the high validity task is learned first. Similarly, subjects who start with the high validity task perform better in a subsequent low validity task than those who start with the low validity task. The authors conclude that if a task of low predictability is to be learned, achievement will be increased by prior exposure to a similar task of high predictability.

Another vein of research has demonstrated that subjects do better if they are given accurate information concerning the meaning of the cues. Miller [1971] has demonstrated substantial differences for tasks in which the cues are unlabeled, falsely labeled or correctly labeled. It was found that, except for a group of professional statisticians, performance was degraded when cues conflicted logically in comparison with cases where there was no logical conflict.

The literature described above gives a general introduction to the methodology and major findings of judgment research within the lens model paradigm. The remainder of this section is devoted to studies specifically aimed at analyzing the effects of more information. The purpose is to describe the current state of knowledge in this area and to demonstrate how the studies reported in this research relate to previous studies.

The first problem involves a more precise definition of the concept of information. For example, an increase in R_e means that a higher proportion of variance may be explained and, thus, the limit on achievement is increased. In this sense, Naylor and Schenck [1968] have demonstrated that subjects can, indeed, take advantage of the increment. Another factor to be considered is the number of cues available. A series of studies on the analysis of test results by clinicians has explored the extent to which additional test data increase accuracy and reliability of judgment. Some representative results from this research will be examined and the reader is referred to Goldberg [1968] for further discussion.

Hunt and Walker [1966] considered the differential diagnosis of schizophrenics, psychoneurotics, organics, retardates and normals on the basis of combined protocols from Weschler Adult Intelligence Scale (WAIS) Vocabulary and Comprehension tests and on the basis of the individual tests alone. They found no improvement in accuracy by the use of combined tests. Golden [1968] employed identifying data, Rorschach, TAT and MMPI tests under conditions of identifying data lone, pairs of tests and all three tests. He found that neither validity nor reliability of judgment was significantly increased as a function of the number of tests used.

A finding which has become almost a classic for its simplicity and ease of interpretation is that of Oskamp [1965]. A total of 32 judges were presented with background case material in four sections. After reading each section, the judge answered a series of questions about the case and estimated the percentage of questions which he answered correctly. At the end of the first section, the average judge was correct on 26 per cent of the questions while he had estimated his accuracy at 33 per cent. After the fourth section, his accuracy was 28 per cent while his estimation was for 53 per cent accuracy. The clear finding is that increasing information did not significantly improve accuracy while confidence increased dramatically. A similar result is reported in Ryback [1967].

A somewhat different finding is supplied by Hammer and Ringel [1965]. Using a simulated military pursuit problem, they presented sequential information about enemy troop movements and asked judges to predict which of several possible targets was being approached. They found that increased information (number of troop movements observed) resulted in higher accuracy and increased confidence.

In a laboratory situation, Brehmer [1973] has discussed the relationship between single-cue and multiple-cue problems. The author points out that for single-cue tasks, judges tend to utilize the cue such that the cue utilization coefficient is higher than the cue validity. On the other hand, in a multiple-cue task, subjects tend to over-match low validity cues and under-match high validity cues. The author conjectures that this discrepancy could be due to: (1) inability of subjects to learn the cue validities in a multiple-cue task; or (2) inability to appropriately integrate information in a multiple cue task. His results suggested that at least part of the suboptimal use of cues in multiple cue studies is due to interference effects.

Dudycha and Naylor [1966] used six two-cue tasks where the first cue validity was either 0.40 or 0.80 and the second cue had validities of 0.20, 0.40 and 0.60; all cues were kept orthogonal to eliminate the effects of cue redundancy. The primary result was that adding a low validity cue to one of high validity decreased performance where the addition of a high validity cue to one of low validity increased performance. A secondary finding was that a probability matching strategy was found to be more descriptive of performance than was an "optimal" strategy.

A fundamental assumption that has guided much of the research into the effects of increased amounts of information is that man has a limited capacity for information processing particularly with regard to multi-dimensional inputs. It is somewhat paradoxical that, as Shepard [1964] has pointed out, man is particularly adept at processing large amounts of data on a perceptual level and yet has difficulty in assimilating the information into a judgment. To quote Shepherd,

> At the level of the perceptual analysis of raw sensory inputs, man evinces a remarkable ability to integrate the responses of a vast number of receptive elements according to exceedingly complex nonlinear rules. Yet once the profusion and welter of this raw input has been thus reduced to a set of usefully invariant conceptual objects, properties and attributes, there is little evidence that they can in turn be juggled and recombined with anything like this facility. On the contrary, the contention that they can belies the obvious disparity between the effortless speed and surety of most perceptual decisions and the painful

> hesitation and doubt characteristic of these subsequent "higher level" decisions [p. 262-263].

Without any doubt, the classic paper concerning man's limited information processing capacity is that of Miller [1956]. In a remarkably readable and far-ranging account, Miller has pointed out the rather severe limitations on the number of perceptual units that man can handle simultaneously. It is findings of this sort which have led to the hypothesis that the human attempts to employ simplifying strategies in order to cope with large information inputs.

In this regard, Bruner, Goodnow and Austin [1956] speak of "the cognitive strain involved in assimilating information" (p. 82). Within the realm of concept formation, they say, "When the nature of a task imposes a high degree of strain on memory and inference, the strategy used for coping with the task will tend to be less conducive to cognitive strain" (p. 112), and later, "When we speak, then, of a strategy as being a move in the direction of efficient or inefficient strain reduction, such a statement must be modified to refer to particular kinds of situations. Some strategies, to be sure, will deal effectively with a wide range of situations; others will be found wanting as the cognitive going gets rough" (p. 113).

The general term "cognitive economy" is proposed by Bruner, Goodnow and Austin although it is not defined. Clearly, however, it refers to the general drive for reduction of cognitive strain. Recently, Hormann [1971] has presented research relating to the design of a large, computer-based system for planning and problem solving which is based on the concept of cognitive economy.

The above research into the use of simplifying strategies in a search for cognitive economy have focused on the qualitative aspects although Bruner, Goodnow and Austin do present empirical evidence. Einhorn [1970, 1971] has formulated an analytical approach to studies of this nature. He hypothesizes three possible models which are candidates for the description of decision making behavior when the judge attempts to reduce cognitive strain in assessing preferences for multi-attributed alternatives.

One such model, the lexicographic model, assumes that the attributes are ordered in importance with lower level attributes being employed for selection only in the case that equal values are obtained on higher level attributes. Although intuitively simple, lexicographic utility models are mathematically complex and hence are not used by Einhorn. It should be pointed out, however, that Fishburn [1968] has stressed that ". . . lexicographic utilities are not necessarily incompatible with numerical utilities."

The second model considered by Einhorn is the conjunctive model which implies that an alternative is acceptable if and only if it surpasses some minimum standard on each attribute. Thus, an unacceptably low level of any attribute is sufficient to render the alternative unacceptable. The important point is that no compensation is allowed since increasing the level of any attribute cannot compensate for an unacceptably low level of another attribute.

The mathematical approximation to the conjunctive model presented by Einhorn [1970, 1971] is

$$U = \prod_{i=1}^{K} x_i^{a_i} \qquad (10)$$

where

U = utility of the multi-attributed distal variable.

x_i = *ith* attribute.

a_i = *ith* parameter value.

Taking logs, Equation (10) can be rewritten:

$$\log U = \sum_{i=1}^{K} a_i \log(x_i). \qquad (11)$$

Clearly, the relationship in (11) fits within the framework of the general linear statistical model and, hence, standard least-squares procedures may be employed to estimate the parameters.

Einhorn also considers a disjunctive model which approximates a situation in which the overall preference for an alternative is based on its highest attribute. In this case, let

$$U = \prod_{i=1}^{K} [1/(a_i - x_i)]^{b_i}, \qquad (12)$$

where

a_i = *ith* constant value,

b_i = *ith* weighting parameter,

and the other quantities are as defined above. Again taking logs,

$$\log U = \sum_{i=1}^{K} - b_i \log(a_i - x_i) \qquad (13)$$

and, once more, least-squares techniques are appropriate.

The models outlined above are obviously somewhat more complex than the standard first-order model mentioned previously. Einhorn [1971] has pointed out, however, that one must be careful to distinguish mathematical from cognitive complexity. While more complex mathematically, the conjunctive and disjunctive models may reasonably be considered to be cognitively simpler. In addition, they represent an attempt to present a specific alternative to standard first-order model rather than a general search for alternative usages which are not based on specific hypotheses.

The tasks used by Einhorn [1970, 1971] involved a job preference decision and a graduate school admissions decision. In general, he found considerable usage of first-order, conjunctive and disjunctive models where, in this context, a subject is said to use a particular model if that model fits his responses better than the alternative models. It was found that the model used was a function of the task involved with particular models being more widely used for certain tasks than for others.

Among the more important findings of Einhorn's work is that the goodness of fit of all models decreased with increasing information. It is not clear whether this lack of fit is due to the use of models fundamentally different from those tested here or due to an increase in the random error. In other words, one cannot conclusively state either that the subjects tended to increasingly use cognitively simpler models, different from the ones examined here, as the number of cues increased or that subjects simply showed an increasing tendency to respond randomly as the number of cues increased. Thus, it is not possible to separate the lack of fit from the pure error.

In summarizing the work of Einhorn [1970, 1971], one must conclude that the particular models considered all tended to fit less well as the number of cues increased, there was evidence of use of the conjunctive and disjunctive as well as of the first-order model (see also, Goldberg [1971]), and there was no clear tendency towards the use of conjunctive and disjunctive models as the number of cues increased. It must be kept in mind that, in these studies, the task was to estimate the overall utility of a multi-attributed alternative. This can be viewed as estimating the utility of a vector variable from its components. It is not clear that findings from a task of this type would also pertain to inference tasks in which the judge uses several cues to predict or infer the value of a distal variable. Since preferences were used, there is no observable criterion value for Einhorn's tasks so that achievement, learning and other measures which require the comparison of Y_s to Y_e cannot be made.

The final study to be examined in this section is the study that corresponds most closely to the research reported here which will be outlined in the next section. Nystedt and Magnusson [1972] asked two groups of judges to predict students' achievement scores on the basis of two, four and six tests. One of the two groups had information about cue validities and intercorrelations. It should be noted that, of course, R_e increased as the number of cues increased. The general findings were: (1) for both groups, achievement increased from two to four cues and decreased from four to six cues; (2) R_s was quite high for all groups and information sets although the group with no cue validity information showed a slight decreasing trend for R_s as the number of cues increased; (3) relative achievement (r_a/R_e) showed the same general pattern as achievement; (4) the group with information about cue validities and cue intercorrelations performed somewhat better than did the other group. For present purposes, the most interesting of findings was that achievement was not a monotonic increasing function of the number of cues.

The studies reviewed in this section do not present an encouraging picture of the ability of humans to make judgmental decisions with increasing amounts of information. To summarize briefly, increasing the amount of information available to the decision-maker does not seem to significantly increase the accuracy or the consistency of judgments. In addition, the judge's policy becomes increasingly more difficult to represent paramorphically as the information level increases, although it is not clear whether this is due to increasing lack of fit for the models used or increasing randomness in the judgments observed.

Outline of the Research

Having introduced the area of judgment research, stated the purpose of the research to be presented here and reviewed the relevant literature, it is now possible to outline the research presented in this report. The general problem is one of determining the effects of the number of cues presented to the judge on inference performance with special attention given to cue validities, cue intercorrelations, overall task predictability and task structure. The studies all involve the inference process, i.e., given a set of cues or predictor variables, the task is to predict or infer the value of some criterion variable. All of the studies involve the use of equally valid cues and all studies are conducted in a laboratory setting under controlled conditions.

Experiment one, which involves increasing the number of cues while holding the overall task predictability constant at each of two levels, is presented in Chapter II. In this case, the cues are correlated, but have the special property that the partial correlations between cues are zero when the criterion variable is held constant. In other words, the only variance shared by cues is criterion variance.

If the number of correlated cues increases with task predictability held constant, the degree of cue intercorrelation must change. Thus, in Chapter III, experiment two is presented which controls the cues to be orthogonal. The experimental design is the same as that for experiment one in all other respects which allows for direct comparisons between the studies.

In Chapter IV the emphasis is on cue utilization. Since cue utilization may be studied in a rather straightforward manner when the cues are uncorrelated, the results of experiment two are re-analyzed in order to search for changing patterns of cue usage as the number of cues is increased. The latter part of the chapter is concerned with maintaining constant cue validities as the number of cues is increased. A third experiment, which examines this case, is described and analyzed.

Experiment four, presented in Chapter V, examines a task that is structurally different from the previous ones. The task is of an hierarchical nature and the analysis centers around the effects of different amounts of structural insight on performance. The working hypothesis is that increased insight into the structure of the task will lead to improved performance.

It should be pointed out that all of the tasks used in this research involve the use of equally valid cues. This is clearly a very special class of inference tasks and the use of such a restricted class of tasks presents certain problems relating to the generalization of the findings. If one is interested in policy capturing research, it is clearly better to employ tasks which do not involve equally weighted cues since such tasks are not diagnostic with respect to the judge's ability to weight cues in an appropriate manner. In the research presented here, however, the emphasis is on performance rather than policy description and the equal weighting case provides a reference point for the assessment of the human capacity to process large numbers of cues.

To summarize, this research consists of a series of four experiments conducted to examine changes in inference behavior as a function of the number of cues provided to the decision maker. The task parameters which are systematically varied include task predictability, task structure, cue validities and cue intercorrelations. Attention is restricted to the case of equally weighted cues. A final chapter reviews the findings of these studies and comments on areas for future research and methdological problems involved in conducting studies involving varying cue set sizes.

CHAPTER II

INFERENCE BEHAVIOR WITH TASK PREDICTABILITY HELD CONSTANT

Introduction to Experiment One

One generally assumes that more information available to the decision maker will result in better decisions. It is, however, possible that the extra information will lead to distraction or even confusion. Thus, there is a trade-off between the beneficial effects of more information and the detrimental effects of distraction. This may be thought of as a conflict between having "more information" and having "more data." The ability of decision makers to process the information effectively and correctly is, of course, an area for empirical investigation. In multiple cue inference tasks, one may vary the amount of information available by manipulating the statistical limit on achievement while independently varying the number of cues or the amount of data.

The purpose of the research presented in this chapter is to determine the effects of increasing the number of cues in a context free inference task with the total task predictability held constant. Since R_e is held constant over all levels of the number of cues factor, changes in performance can be viewed as being due primarily to the increased strain of handling a larger number of cues. It is possible, however, that changes which occur at one fixed level of R_e may differ from changes at a different fixed level of R_e. In terms of an analysis of variance, this would correspond to a significant number of cues by task predictability interaction.

Unfortunately, it is not possible to completely isolate the number of cues and R_e as quantities to be manipulated without consideration of other factors. In general, if R_e is held constant while the number of cues is varied, the cue validities and/or cue intercorrelations will change. Two cases are immediately suggested for investigation. In the experiment reported in this chapter, the cues are correlated and hence both intercorrelations and validities vary as the number of cues varies.

Table 1. Theoretical Cue Intercorrelations (r_{ij}) and Validities (r_{ie}) for Experiment One

	$R_e^2 = 0.50$		$R_e^2 = 0.80$	
	r_{ij}	r_{ie}	r_{ij}	r_{ie}
2 Cues	0.33	0.58	0.67	0.82
5 Cues	0.17	0.41	0.45	0.67
8 Cues	0.11	0.33	0.35	0.59

The case of uncorrelated cues is examined in experiment two which is discussed in the next chapter. In that situation, of course, only the validities change as a function of the number of cues.

The cue sets for experiment one were generated by considering each cue to be generated as the sum of the value of the criterion variable and an independent error term. In particular,

$$X_j = Y_e + \epsilon, \quad j = 1,\ldots,k$$

where

$$Y_e \sim N(500, 100^2)$$

$$\epsilon \sim N(0, \sigma_\epsilon^2)$$

$$X_j \sim N(500, 100^2 + \sigma_\epsilon^2)$$

$$\mathrm{CORR}(Y_e, \epsilon) = 0.$$

That is, for a given trial, the value of Y_e was drawn from a normal distribution with mean 500 and variance $100^2 + \sigma_\epsilon^2$; then, each of the k cue values was generated as the sum of Y_e and an independent error term which was drawn from a distribution with mean 0 and variance σ_ϵ^2. An independent realization of the random variable was used for each of the k cue values. For each subsequent trial, the procedure was repeated on the basis of a new realization of Y_e. Of course, the parameter values given here refer to population values and the sample values actually used will differ somewhat. Appendix B contains a summary of the observed values of the various quantities involved in experiment one.

A word of explanation regarding the method of data generation is in order. It is clear that individual cues are intercorrelated since they share the variance of the criterion variable, Y_e. They do, however, have the special property that their intercorrelations are zero (within sampling error) when the variance of Y_e is partialled out. This form of data may arise in practice when each cue is an unbiased but unreliable measure of the criterion. No claim is made, however, regarding the correspondence of this experiment to any particular real-world task.

It is clear that the variances of Y_e and ϵ along with their independence determine the correlations among the cues and the cue validities. These, in turn, determine the multiple correlation coefficient, R_e. This suggests that by appropriate selection of the variances, one may establish some desired value of R_e and since in this experiment the variance of Y_e was kept constant, it was possible to systematically vary the number of cues and R_e is a factorial fashion by appropriately choosing σ_ϵ^2.

In experiment one, two levels of R_e were used: $R_e^2 = 0.50$ and $R_e^2 = 0.80$. Two, five, and eight cues were employed with the number of cues factor and the R_e factor varied in a factorial arrangement. This is schematically represented in Figure 2.

		Number of Cues		
		2	5	8
Level of R_e^2	0.50			
	0.80			

Figure 2. Experimental Design for Experiment One

In addition, 3 blocks of 50 trials each were employed for each condition so that the final design was a 2(levels of R_e) × 3(number of cues) × 3(blocks of trials) factorial with repeated measures on the last factor. This design is discussed and analyzed in detail by Winer [1961]. The quantities analyzed were n_a, G and R_s. The analysis was performed on transformed variables obtained by Fisher's Z-transformation which renders the sampling distribution of the correlation coefficient approximately normal with variance $\frac{1}{N - 3}$ when sampling from a bivariate normal population with a true correlation of zero. The Z-transformation is not strictly appropriate for multiple correlation coefficients such as R_s and G; however, it has been used in the literature with seemingly acceptable results and, hence, will be used here.

Procedure

Subjects were 30 undergraduate engineering students taking a course in industrial engineering. They were not paid and participated on a voluntary basis. All subjects were male.

An answer sheet with attached instructions was distributed to the subjects. The task was described as one of predicting a number on the basis of a group of other numbers. Subjects were told that perfect accuracy was not possible and that the number to be predicted was an integer between 0 and 1000. Questions were then answered. The detailed instructions are given in Appendix C.

An apparatus was constructed so that the cues could be projected trial by trial onto a screen by means of an overhead transparency projector. For each trial: (1) the trial number and the values of the k cues were presented; (2) subjects recorded their estimate; (3) the value of Y_e was shown and subjects recorded it. Each trial required approximately 20 seconds. One hundred and fifty trials were run with a brief (two or three minutes) pause after each block of 50 trials. The entire process required approximately 1 hour and 6 groups of 5 subjects each were run corresponding to the 6 cells of Figure 2.

After the data had been collected, it was keypunched and analyzed by means of a computer program written to perform analyses associated with the lens model equation. This consists of developing a regression equation of Y_e on the cues for each block of trials. For each subject and each block of trials a regression of Y_s on the cues was calculated.

One further methodological note is in order. Since a regression equation was obtained for each block of 50 trials, the ratio of observations to parameters was quite low, particularly for the case of eight cues. To overcome this, one might use step-wise regression and weight fewer variables or base the regression model on the entire set of trials as was done by Dudycha and Naylor (1966). Since, however, the purpose was to describe the subject's responses in a given block of trials rather than to generate stable regression weights to be used for prediction, these alternatives were rejected and the procedure described above was followed.

Analysis

The ANOVA for achievement is summarized in Table 2. The effect of task predictability, R_e, was highly significant ($p<.0005$); reference to Figure 3 and the data which are given in Appendix D demonstrate that achievement was much higher for the high predictability task than for the lower one. The ANOVA also shows that while the number of cues effect was not significant, there was a marginally significant ($p<.10$) interaction between task predictability and number of cues.[1] It was thus necessary to examine the effect of increasing the number of cues at each level of R_e. To accomplish this, one-way ANOVAs were run on the number of cues

---

[1]The total variation is broken into Between Subjects and Within Subjects components. Within each division, the bottom source is used as the denominator in the relevant F-ratios.

factor for each of the levels of R_e. Table 3 summarizes the analyses and shows that the effect was not significant at the low level of R_e but was at the high level.

Table 2. ANOVA for r_a (Experiment One)

Source of Variation	SS	DF	MS	F
Between Subjects	9.56	29		
A (R_e)	5.39	1	5.39	45.16 ($p<.0005$)
B (Number of Cues)	.51	2	.25	2.12
AB	.80	2	.40	3.33 ($p<.10$)
Subjects Within Groups	2.87	24	.12	
Within Subjects	2.23	60		
C (Blocks of Trials)	.72	2	.36	14.63 ($p<.0005$)
AC	.04	2	.02	<1
BC	.18	4	.05	1.83
ABC	.10	4	.02	1.01
C x (Subjects Within Groups)	1.18	48	.02	

Table 3. ANOVA for r_a at $R_e^2 = 0.50$ and $R_e^2 = 0.80$

	$R_e^2 = 0.50$				$R_e^2 = 0.80$			
Source of Variation	SS	DF	MS	F	SS	DF	MS	F
Number of Cues	.07	2	.03	1.34	.16	2	.08	8.46 ($p<.01$)
Error	.30	12	.03		.12	12	.01	
Total	.37	14			.28	14		

Returning to Table 2, it is seen that achievement did vary significantly over blocks of trials. Figure 4 shows that r_a did not vary monotonically but, rather, increased from the first block to the second and decreased from the second to the third at each level of R_e. It is interesting to note the lack of a significant block of trials by task predictability interaction. This implies that the shapes of the learning curves for achievement did not differ significantly for high versus low predictability tasks.

The analysis of variance for R_s, presented in Table 4, is strikingly similar to that for r_a.

Table 4. ANOVA for R_s (Experiment One)

Source of Variation	SS	DF	MS	F
Between Subjects	13.75	29		
A (R_e)	3.31	1	3.31	10.37 ($p<.005$)
B (Number of Cues)	.81	2	.41	1.27
AB	1.96	2	.98	3.06 ($p<.10$)
Subjects Within Groups	7.67	24	.32	
Within Subjects	3.41	60		
C (Blocks of Trials)	.85	2	.43	8.67 ($p<.001$)
AC	.02	2	.01	<1
BC	.09	4	.02	<1
ABC	.09	4	.02	<1
C × Subjects Within Groups	2.35	48	.05	

Once again, there is a significant task predictability effect, a weak interaction between R_e and the number of cues and a strong effect for blocks of trials. The significant interaction necessitates an analysis of the number of cues factor at each level of R_e; these ANOVAs are given in Table 5. The number of cues is found to significantly affect R_s at the high level of predictability but not at the low level.

Table 5. ANOVA for R_s at $R_e^2 = 0.50$ and $R_e^2 = 0.80$

	$R_e^2 = 0.50$				$R_e^2 = 0.80$			
Source of Variation	SS	DF	MS	F	SS	DF	MS	F
Number of Cues	.04	2	.02	<1	.07	2	.04	4.04 ($p<.05$)
Error	.35	12	.03		.11	12	.01	
Total	.39	14			.18	14		

Figures 3 and 4 demonstrate these effects. In particular, it is of interest to note that the r_a and R_s curves have highly similar shapes at each level of R_e. This is true over numbers of cues and blocks of trials. The fit of the first order model to the subjects' responses, measured by R_s, increased from the two cue condition to the five cue condition.

Figures 3 and 4 and the ANOVA of Table 6 indicate that G was significantly affected by both task predictability and number of cues and that there was a marginally significant ($p<.10$) effect of blocks of trials. From Figure 3, G decreases monotonically as the number of cues increases and is lower in the low predictability case than in the high predictability case for the five and eight cue conditions. Figure 4 shows that whereas r_a and R_s had increased from the first to the second block of trials and decreased from the second to the third block,

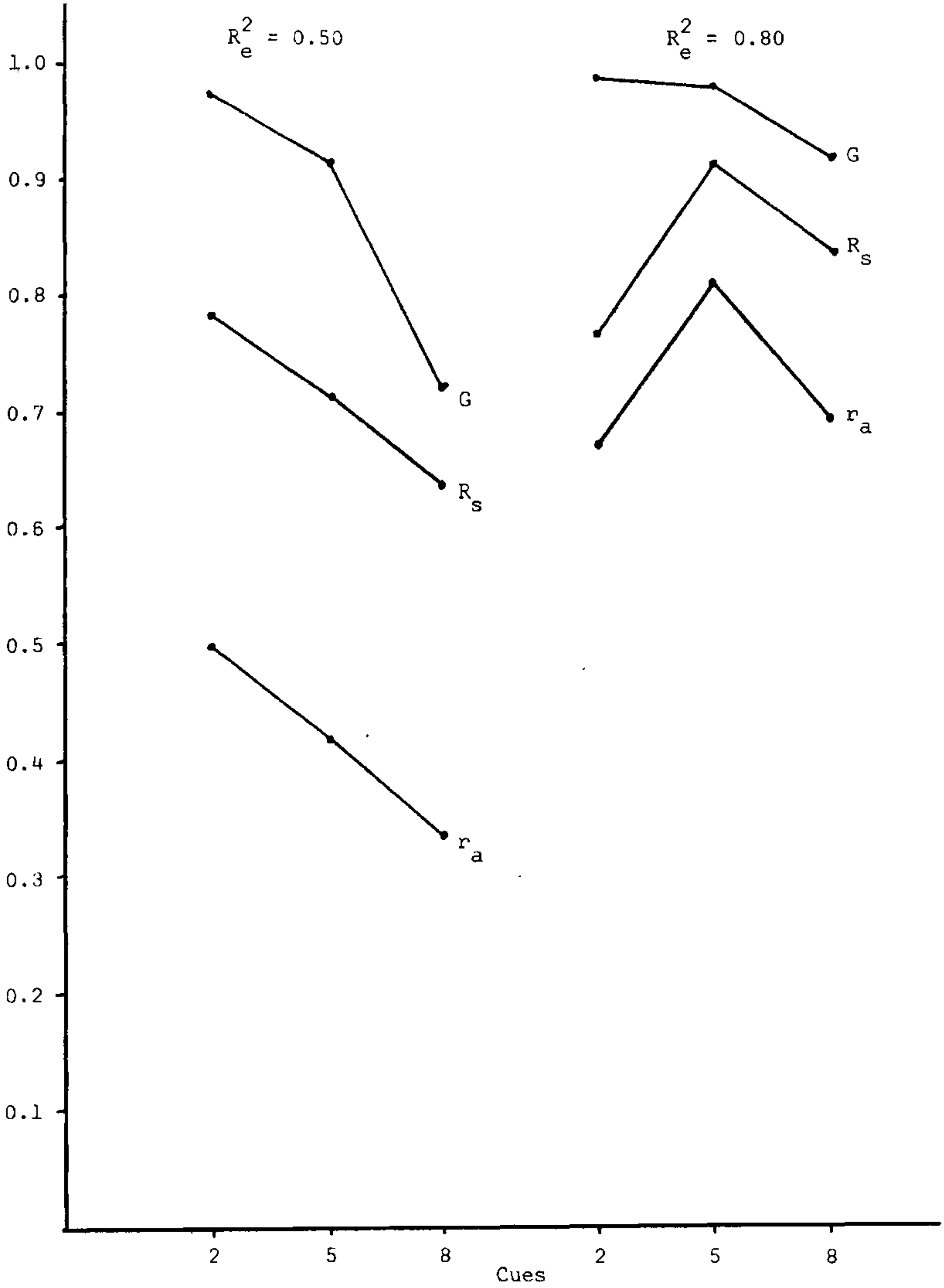


Figure 3. Performance as a Function of Number of Cues (Experiment One)

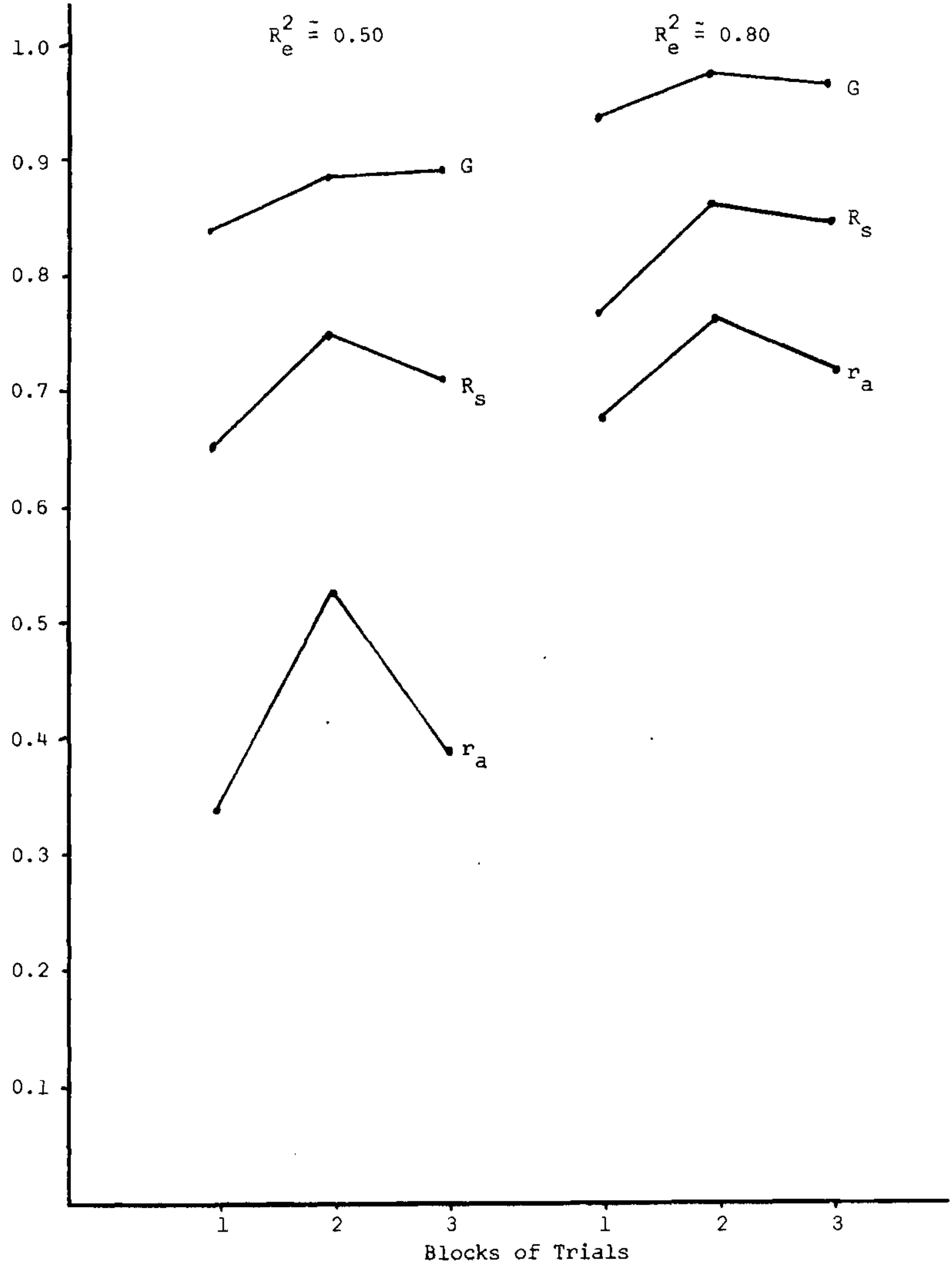


Figure 4. Performance Over Blocks of Trials (Experiment One)

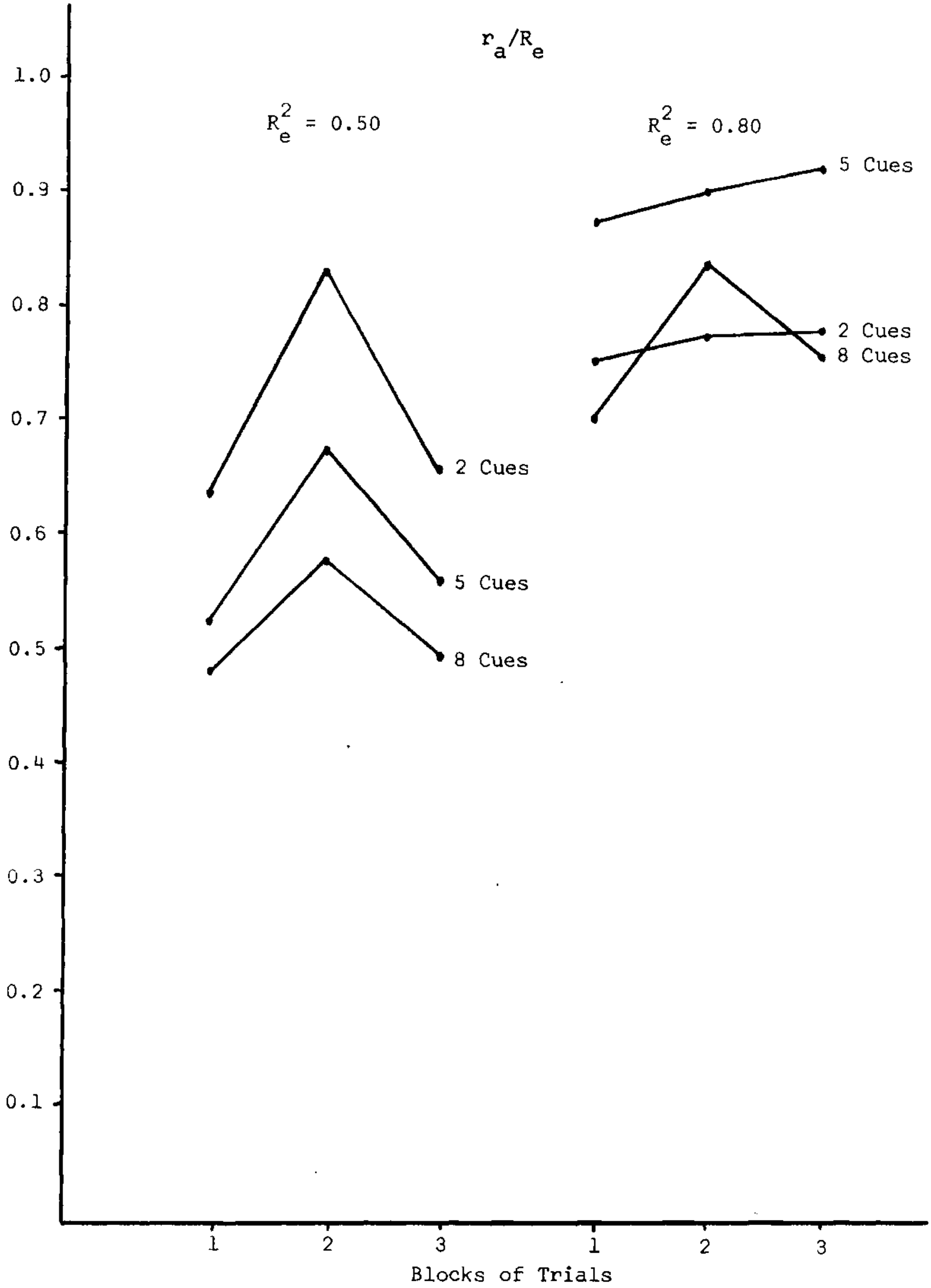


Figure 5. Relative Achievement for Experiment One

G shows a slight increasing trend across the three blocks.

Table 6. ANOVA for G (Experiment One)

Source of Variation	SS	DF	MS	F
Between Subjects	57.01	29		
A (R_e)	8.60	1	8.60	12.28 ($p<.005$)
B (Number of Cues)	30.26	2	15.13	21.62 ($p<.0005$)
AB	1.35	2	.68	<1
Subjects Within Groups	16.80	24	.70	
Within Subjects	25.99	60		
C (Blocks of Trials)	2.03	2	1.02	2.47 ($p<.10$)
AC	1.60	2	.80	1.94
BC	1.02	4	.26	<1
ABC	1.58	4	.40	<1
C x Subjects Within Groups	19.76	48	.41	

An additional analysis was conducted on relative achievement which which is defined as r_a/R_e and is interpreted as achievement as a percentage of maximal possible achievement. Since R_e is subject to sampling error, the empirical values of R_e were used in the analysis. Figure 5 shows relative achievement over blocks of trials for each level of R_e at two, five and eight cues. Notice that at the low level of R_e, relative achievement shows the same pattern across blocks of trials for each number of cues with the two cue condition yielding higher relative achievement than the five cue condition which in turn is higher than the eight cue condition. At the higher level of R_e, the pattern is not as clear although the five cue condition is higher than the others. In general, relative achievement appears to be higher at $R_e^2 = 0.80$ than at $R_e^2 = 0.50$.

To test these results, Kruskal-Wallis nonparametric one-way ANOVAs (see Hollander and Wolfe [1973], p. 114) were conducted on the relative achievement values for each subject across blocks of trials. The null hypothesis to be tested was that relative achievement did not differ as the number of cues was varied; a separate ANOVA was run at each level of R_e and the results are given in Table 7.

Table 7. Kruskel-Wallis One-Way ANOVA on r_a/R_e at Each Level of R_e^2

$R_e^2 = 0.50$ (Ranks in Parentheses)			$R_e^2 = 0.80$ (Ranks in Parentheses)		
2 Cues	5 Cues	8 Cues	2 Cues	5 Cues	8 Cues
.700405(11)	.750477(14)	.485885(5)	.914201(15)	.899046(13)	.859532(9)
.476394(4)	.707073(12)	.212801(2)	.763840(7)	.904763(14)	.817029(8)
.725836(13)	.680941(10)	.643800(8)	.752001(5)	.885117(11)	.589829(2)
.500440(6)	.187948(1)	.227660(3)	.609593(3)	.891390(12)	.680068(4)
.949632(15)	.502878(7)	.650296(9)	.524036(1)	.873616(10)	.757827(6)
$R_1 = 49$; $R_2 = 44$; $R_3 = 27$; H = 2.66 (Not Significant)			$R_1 = 31$; $R_2 = 60$; $R_3 = 29$ H = 6.02 ($p<.05$)		

It should be noted that the effect was significant ($p<.05$) at the high level of R_e but non-significant at the low level. This result, of course, is the same as that obtained for r_a as described in Table 3.

Note, however, that since empirical values of R_e were used in computing relative achievement, the sampling error involved is included in the analyses of Table 7. The similar results of Tables 3 and 7 are thus reassuring.

To test the hypothesis that relative achievement did not differ as a function of task predictability (R_e), the non-parametric Wilcoxon rank sum test large sample approximation was utilized (see Hollander and Wolfe [1973], p. 68). The results, which are summarized in Table 8, show that relative achievement is significantly higher ($p<.002$) at $R_e^2 = 0.80$ than at $R_e^2 = 0.50$.

Table 8. Wilcoxon Rank Sum Test (Large Sample Approximation) on r_a/R_e

$R_e^2 = 0.50$	$R_e^2 = 0.80$
.187948 (1)	.524036 (8)
.212801 (2)	.589829 (9)
.227660 (3)	.609593 (10)
.476394 (4)	.680068 (13)
.485885 (5)	.752001 (19)
.500440 (6)	.757827 (20)
.502878 (7)	.763840 (21)
.643800 (11)	.817029 (22)
.650296 (12)	.859532 (23)
.680941 (14)	.873616 (24)
.700405 (15)	.885117 (25)
.707073 (16)	.891390 (26)
.725836 (17)	.899046 (27)
.750477 (18)	.904763 (28)
.949632 (30)	.914201 (29)
$W^* = 2.97$ ($p<.002$)	

Summary and Discussion of Experiment One

The task employed was of a somewhat special nature in that each cue and the criterion variable all had approximately the same mean and the cues were of equal validity (within sampling error). It is intuitively plausible, and informal post-experiment discussions indicate, that this type of problem suggests an averaging strategy. It should be pointed out, however, that any equal weighting scheme will maximize r_a so that a subject who responds with the average value of the cues will attain the same level of achievement as will one who responds with the sum of the cues. Any deviation from an equal weighting scheme will, however, result in a decrement in achievement.

The general hypothesis was that since R_e was held constant as the number of cues was increased, achievement would be a decreasing function of the number of cues. The results, however, do not support this hypothesis in that it was found that the effect on achievement of increasing the number of cues depends on the level of task predictability, R_e. For the low predictability task (approximately 50 per cent of the variance predictable), increasing the number of cues did not significantly affect achievement; for a high predictability task (approximately 80 per cent of the variance predictable), a significant effect was noted but it was not monotonic as achievement increased from two to five cues and decreased from five to eight cues.

This result confirms and extends the findings of Nystedt and Magnusson [1972] who noted an increase in achievement from a two cue to a four cue problem and a subsequent decrease from the four cue problem to one involving six cues. In their study, however, R_e increased with the number of cues and, thus, the effect of task predictability was confounded with the effect of the number of cues.

In the present research, it was hypothesized that the fit of a first-order model would decrease with the number of cues. An analysis of the fit of the model, as measured by R_s, indicates, however, that R_s follows to a remarkable extent the pattern of r_a. Thus, for $R_e^2 = 0.50$, the number of cues did not significantly affect R_s whereas for $R_e^2 = 0.80$, R_s was affected but in the non-monotonic fashion described above for r_a.

G, a measure of the extent to which the portion of the response system which can be described by a first-order model is proportional to the first-order environmental model, was significantly affected by R_e and by the number of cues with no interaction between these factors. G decreased monotonically as a function of the number of cues and was lower for the low predictability task than for the high predictability task.

Over blocks of trials, r_a and R_s varied significantly while G displayed a marginally significant ($p<.10$) effect. The blocks of trials factor did not significantly interact with any other factor for r_a, G or R_s.

The results of this experiment are nicely interpretable in terms of Hammond and Summers' [1972] concept of cognitive control. Briefly stated, this theory hypothesizes that cognitive skills are composed of two factors--knowledge, the detection of the appropriate policy as measured by G, and cognitive control, the ability to consistently utilize this policy, as measured by R_s. Since both G and R_s must approach unity for achievement to approach its upper limit and since G and R_s are statistically independent, knowledge and cognitive control may be thought of as independent components of achievement.

Consider Figure 4 which shows r_a, G and R_s as a function of the blocks of trials. For the $R_e^2 = 0.50$ condition, achievement increased from the first to the second block and decreased from the second to the third block. Presumably, this represents a fatigue or boredom effect after 100 trials. Notice, however, that G remains high even in the third block indicating, with respect to the theory of cognitive control, that control decreased while knowledge remained high. In other words, the subjects retained a knowledge of the correct policy but were unable to consistently employ it, presumably due to fatigue or boredom.

It must be noted that the concept of cognitive control is somewhat conjectural. The first-order model of the judge is "paramorphic" and is not an indication of the underlying cognitive processes. Thus, the interpretation given above must be considered in this light.

CHAPTER III

THE CASE OF UNCORRELATED CUES

Introduction to Experiment Two

In the previous chapter, performance in a multiple-cue inference task was examined as a function of the number of cues. The findings indicated that achievement, r_a, is not a simple, monotonic function of the number of cues but rather that the effect of adding cues depends on the level of task predictability, R_e. It was pointed out that when the cues are correlated, the cue validities and cue intercorrelations must vary from condition to condition if R_e is to be held constant as the number of cues is varied. The cues for experiment one were developed by generating a distribution of Y_e values and then computing each cue, X_i, as the sum of Y_e and an independent random error term.

Experiment two was designed to obviate the effect of varying cue intercorrelations in such a task by restricting the cues to be uncorrelated. This was accomplished by generating the cue values first and then computing Y_e as a linear combination of the X_i and an independent error term.[1] In particular:

---

[1]Connolly [1974] distinguishes between "cues," which share criterion variance as in experiment one, and "components," which do not share criterion variance, and, thus, may be uncorrelated. He contends that there is a fundamental distinction between cues and components which he discusses in detail.

$$Y_e = \sum_{i=1}^{k} X_i + b + \epsilon$$

where

$$X_i \sim N(\mu_x, \sigma_x^2)$$

$$\epsilon \sim N(0, \sigma_\epsilon^2)$$

$$Y_e \sim N(k\mu_x + b,\ k\sigma_x^2 + \sigma_\epsilon^2)$$

$$X_i \text{ and } \epsilon \text{ independent for all } 1 \le i \le k$$

$$X_i \text{ and } X_j \text{ independent for all } i \ne j.$$

The parameters σ_x^2, σ_ϵ^2 and b were chosen for each condition so that Y_e had mean 500 and variance 50^2. In addition, each cue had mean 500 in the population. Table 9 gives the resulting (theoretical) cue intercorrelations (r_{ij}) and validities (r_{ie}) for experiments one and two. Appendix E contains a comparison of the theoretical and empirical values of the task parameters for experiment two.

There are at least three reasons for considering the case of uncorrelated cues. In the first place, this eliminates the confounding of a cue intercorrelation effect with the number of cues effect. It must be pointed out, of course, that the cue validities still do vary with the number of cues. A second reason for an interest in orthogonal cues is that in this case the cue utilization coefficients (r_{is}) are meaningful measures of the extent to which a particular cue or subset of cues is utilized.

Table 9. Theoretical Cue Intercorrelations (r_{ij}) and Validities (r_{ie}) for Experiments One and Two

	EXPERIMENT ONE				EXPERIMENT TWO			
	$R_e^2 = 0.50$		$R_e^2 = 0.80$		$R_e^2 = 0.50$		$R_e^2 = 0.80$	
	r_{ij}	r_{ie}	r_{ij}	r_{ie}	r_{ij}	r_{ie}	r_{ij}	r_{ie}
2 Cues	0.33	0.58	0.67	0.82	0.00	0.50	0.00	0.62
5 Cues	0.17	0.41	0.45	0.67	0.00	0.32	0.00	0.40
8 Cues	0.11	0.33	0.35	0.59	0.00	0.25	0.00	0.32

The reader is referred to Hoffman (1960,1968) and Darlington (1968) for discussions of measures of cue utilization. Finally, there is evidence (see Naylor and Schenck (1968)) that performance in a multiple cue inference task is influenced by the degree of cue intercorrelation. Thus, both correlated and uncorrelated cues are of interest from a behavioral standpoint.

The findings of Naylor and Schenck (1968) are of particular significance for a comparison of experiments one and two. Using two-cue inference tasks, they found that achievement (r_a), relative achievement (r_a/r_e) and response linearity (R_s) were all higher for higher levels of cue intercorrelation (r_{12}) and task predictability (R_e). It was also found that the facilitating effect of increasing redundancy was greater at high levels of R_e. Thus it was hypothesized that performance in experiment one would be higher than performance in a corresponding condition in experiment two and that the performance differences would tend to be greater for tasks for high predictability (R_e^2=0.80) than for tasks of low predictability (R_e^2=0.50). It was expected that this would be true for the cases of two, five and eight cues.

One note of caution is appropriate with respect to a comparison of experiments one and two. Schenck and Naylor (1968) have shown that for a given set of cue weights, any increase in cue intercorrelation will necessarily lead to an increase in R_s. Thus, the findings for R_s must be viewed with caution.

Procedure

The subjects for experiment two were 30 undergraduate students taking a course in industrial engineering. Twenty-seven subjects were male and three were female. All participated on a purely voluntary basis receiving neither money nor course credit for their participation. The procedure was identical to that used in experiment one as described in Chapter II.

The experimental design, as in experiment one, was a 2 (levels of R_e) × 3 (number of cues) × 3 (blocks of trials) factorial with repeated measures on the last factor. The basic data analysis consisted of ANOVAs on r_a, G and R_s with Fisher's Z-transformation employed. These results and subsequent analyses are discussed in the next section.

Analysis

The response data for experiment two are given in Appendix F. The results are summarized in Figure 6 for each level of task predictability and each number of cues.

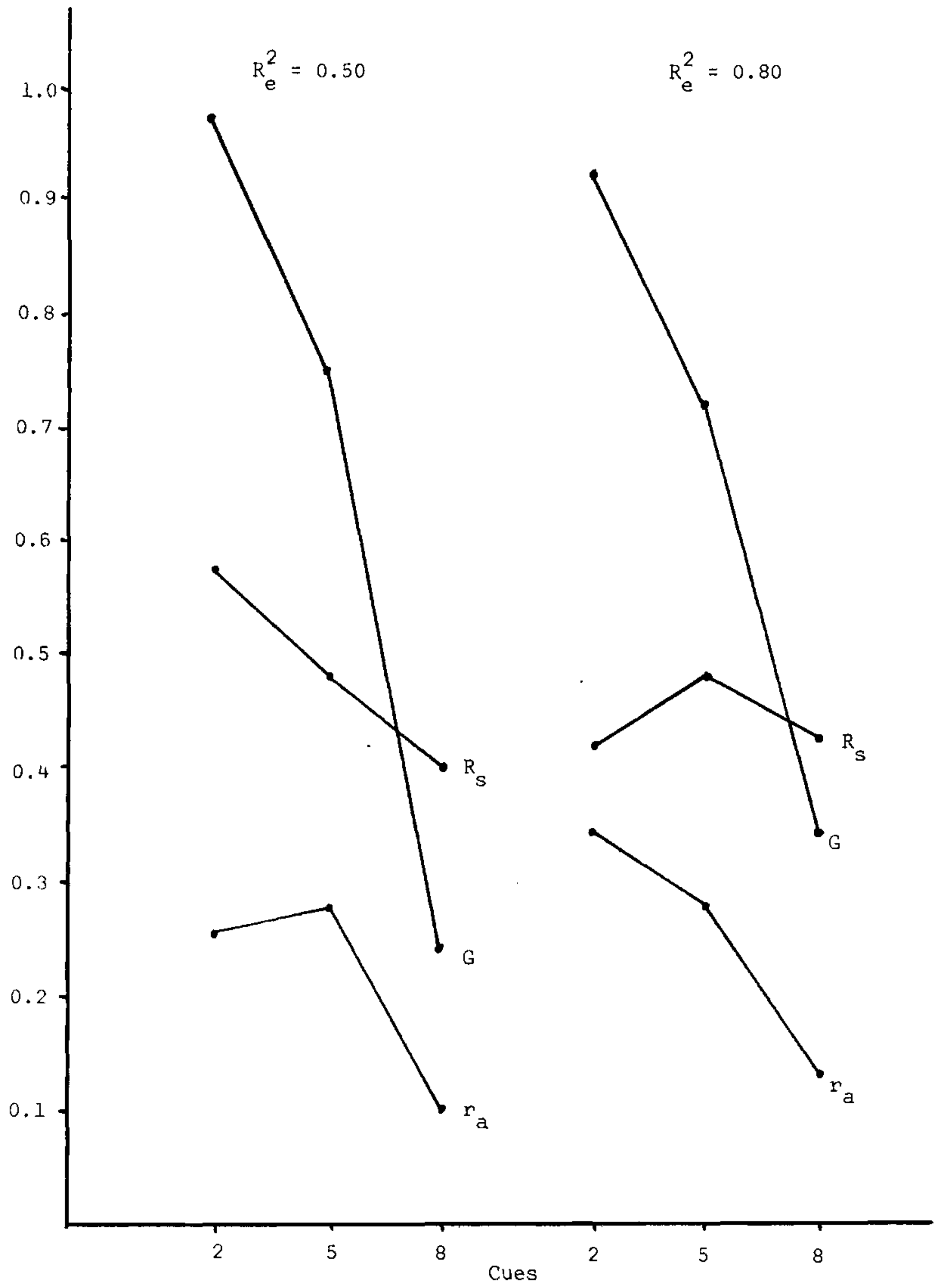


Figure 6. Performance as a Function of Number of Cues (Experiment Two)

The analysis of variance for r_a is given in Table 10. It is seen that the effect for number of cues is highly significant. This effect is shown in Figure 6 where it is seen that there is a general decreasing trend in r_a as the number of cues increases with the exception of going from two to five cues at $R_e^2 = 0.50$. The shapes of the two r_a curves in Figure 6 do not differ significantly as the level of R_e by number of cues interaction does not approach significance.

Table 10. ANOVA for r_a (Experiment Two)

Source of Variation	SS	DF	MS	F
Between Subjects	2.06	29		
A (R_e)	.03	1	.03	<1
B (Number of Cues)	.76	2	.38	7.25 (p<.005)
AB	.01	2	.01	<1
Subjects Within Groups	1.27	24	.05	
Within Subjects				
C (Blocks of Trials)	.22	2	.11	3.50 (p<.05)
AC	.03	2	.01	<1
BC	.05	4	.01	<1
ABC	.37	4	.09	3.04 (p<.10)
C x (Subjects Within Groups)	1.48	48	.03	

It is of considerable interest to note that R_e did not significantly affect achievement. This is true even for the case of two cues ($t=1.29$, $df=8$, $p<.25$). There is a significant blocks of trials effect which is portrayed in Figure 7. In addition, there is some evidence of

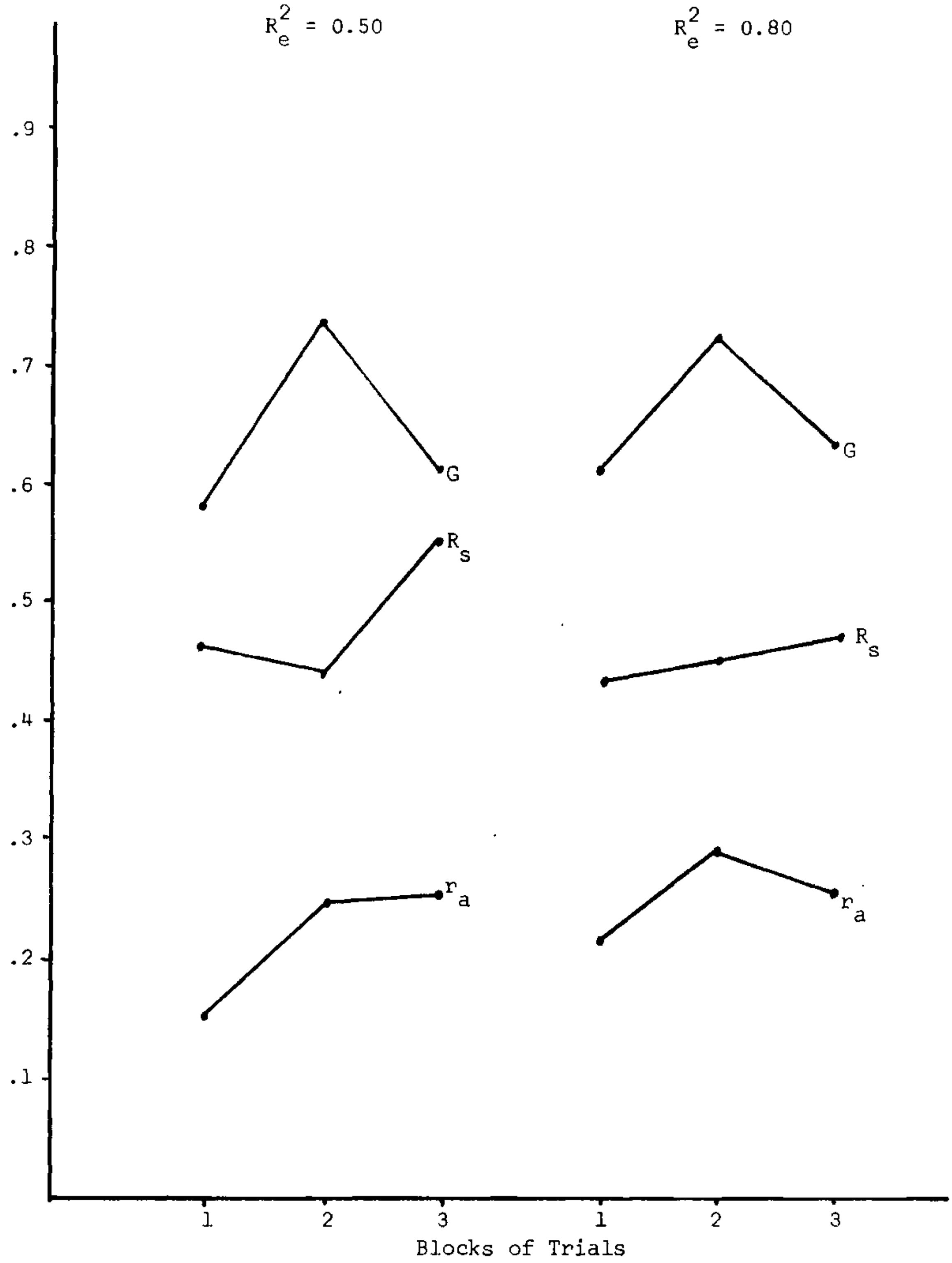


Figure 7. Performance Over Blocks of Trials (Experiment Two)

an R_e by number of cues by blocks of trials interaction but since it is neither strong ($p<.10$) nor readily interpretable, it will not be pursued here.

Table 11 gives the analysis of variance for R_s. It will be noted that no main effects or interactions are significant. In particular, the number of cues does not significantly decrease the fit of the first order model. This finding does not coincide with the findings of Einhorn (1971). Since he used much different tasks, as has been discussed above, such a discrepancy is not surprising.

Table 11. ANOVA for R_s (Experiment Two)

Source of Variation	SS	DF	MS	F
Between Subjects	2.41	29		
A (R_e)	.07	1	.07	1.00
B (Number of Cues)	.26	2	.13	1.76
AB	.32	2	.16	2.20
Subjects Within Blocks	1.76	24	.07	
Within Subjects	8.13	60		
C (Blocks of Trials)	.24	2	.12	<1
AC	.12	2	.06	<1
BC	.47	4	.12	<1
ABC	.28	4	.07	<1
C x (Subjects Within Groups)	7.02	48	.15	

The ANOVA for the matching index, G, is summarized in Table 12.

Table 12. ANOVA for G (Experiment Two)

Source of Variation	SS	DF	MS	F
Between Subjects	70.49	29		
A (R_e)	.60	1	.60	<1
B (Number of Cues)	40.39	2	20.19	17.58 ($p<.0005$)
AB	1.94	2	.97	<1
Subjects Within Groups	27.56	24	1.15	
Within Subjects	40.02	60		
C (Blocks of Trials)	1.24	2	.62	1.00
AB	1.51	2	.75	1.21
BC	.76	4	.19	<1
ABC	6.70	4	1.67	2.70
C x (Subjects Within Groups)	29.81	48	.62	

It is seen that the only significant effect is that of the number of cues. Reference to Figure 6 shows that increases in the number of cues lead to rather dramatic decreases in G. Since the optimal policy involves weighting the cues equally, and since G is a measure of the "match" between the optimal policy and the actual policy, G may be interpreted as a measure of the extent to which an equal weighting scheme is employed. In this case, it seems quite clear that increasing the number of uncorrelated cues leads to a departure from equal weighting. This implies that some cues are weighted more heavily than others and lends support to the hypothesis that one way in which subjects simplify tasks involving large cue sets is to utilize a subset of the cues. This suggestion is discussed in more detail in the next chapter.

Discussion and Comparison with Experiment One

It was hypothesized earlier in this chapter that in comparing experiment one (correlated cues) with experiment two (uncorrelated cues) one would find that achievement would tend to be lower with uncorrelated cues than under corresponding conditions with correlated cues. Furthermore, the difference in achievement would be greater at $R_e^2 = 0.80$ than at $R_e^2 = 0.50$. Comparison of Figures 3 and 6 lends subjective support to this conjecture. In order to test the hypothesis statistically one might consider performing a three-way ANOVA where the factors are task predictability, number of cues and degree of cue intercorrelation. It has been pointed out previously, however, that cue intercorrelation varies as a function of number of cues and task predictability. Thus, the resulting design would not constitute a factorial design. For this reason, t-tests were conducted on r_a between the correlated and uncorrelated cases for each number of cues and each level of R_e. The null hypothesis to be tested is that the difference in mean achievement for the two conditions is zero. The tests were conducted on the Z-transformed data. The results are given in Table 13.

Table 13. Tests on Mean Achievement Differences for Correlated Versus Uncorrelated Cues (DF=8)

	$R_e^2 = 0.50$	$R_e^2 = 0.80$
2 Cues	t=2.91 (p<.02)	t=3.75 (p<.006)
5 Cues	t=1.22 (NS)	t=13.89 (p<.00001)
8 Cues	t=2.54 (p<.04)	t=6.94 (p<.0002)

The use of several t-tests as described above leads to a distortion of the nominal significance level. In particular, for a given level of significance, the probability of obtaining at least one significant difference increases with the number of individual tests performed. This analysis should thus be interpreted with some caution.

The findings here clearly support the hypothesis stated above which was based on the findings of Naylor and Schenck (1968). There is a general trend towards higher achievement with correlated cues than with uncorrelated cues and this trend is much more dramatic at the higher level of R_e^2. In addition, it is seen that this result can be extended from the two-cue case studied by Naylor and Schenck to the larger cue sets employed here.

One rather striking difference between experiments one and two is that in the former case there was a highly significant ($p<.0005$) task predictability effect which was not present in the latter case. Since no known previous study has employed large numbers of uncorrelated cues, there is no comparison to be made with previous results. As was previously mentioned, there is no significant difference even for the two-cue case ($t=1.29$, $df=8$, $p<.25$) though the trend is in the predicted direction. This result apparently conflicts with the Naylor and Schenck (1968) finding. Their study, however, found a significant R_e effect over several levels of cue intercorrelation so that the comparison is not direct.

The comparison of the fit of the first-order model, as measured by R_s, also shows differences between the two experiments. With correlated cues, the pattern is similar to that for achievement with R_s significantly higher for the $R_e^2 = 0.80$ condition than for the $R_e^2 = 0.50$ condition. This is definitely not the case for uncorrelated cues. Also, in the second study, there was no significant number of cues effect whereas experiment one yielded a significant task predictability by number of cues interaction and a significant number of cues main effect at $R_e^2 = 0.80$.

The matching index, G, shows a similar monotonic decreasing pattern as a function of the number of cues in each study. In experiment one, however, G remained at relatively high level even for eight cues while in experiment two there is a dramatic decrease in G as the number of cues increases. In the second study, G is unaffected by task predictability while in the first experiment, G is significantly higher at $R_e^2 = 0.80$ than at $R_e^2 = 0.50$.

To summarize, it appears that the relationship of achievement, r_a, response linearity, R_s, and matching, G, to the number of cues is complex, depending at least on the level of task predictability and the degree of cue intercorrelation. With uncorrelated cues, there is a general decreasing trend in r_a and G as the number of cues increases with task predictability having little or no influence. When cue intercorrelations are positive, the pattern is much more complicated with the effects of increasing the number of cues depending on the level of R_e^2. With orthogonal cues, G is apparently the major determinant of achievement whereas with positive cue intercorrelations, R_s becomes more important. Evidently, when the cues are correlated, subjects find it easier to generate appropriate strategies. It should be noted again that Naylor and Schenck [1968] have demonstrated that R_s will necessarily increase as cue redundancy increases, provided the cue weights remain constant.

CHAPTER IV

CUE UTILIZATION AS A FUNCTION OF THE NUMBER OF UNCORRELATED CUES

Usefulness of Uncorrelated Cues

It was pointed out in the previous chapter that, in general, the manipulation of the number of cues and R_e^2 will lead to varying degrees of cue intercorrelation. Thus, experiment two was conducted to control for this phenomenon by restricting the cues to be uncorrelated. It is also true that when the cues are correlated, there is no generally acceptable measure of cue importance. For example, consider a two cue problem with regression coefficients:

$$\beta_1 = \frac{r_{1e} - r_{2e}r_{12}}{1 - r_{12}^2} \frac{\sigma_{y_e}}{\sigma_{x_1}}$$

$$\beta_2 = \frac{r_{2e} - r_{1e}r_{12}}{1 - r_{12}^2} \frac{\sigma_{y_e}}{\sigma_{x_2}}$$

Clearly, each coefficient is a function not only of the validity of the corresponding independent variable, but also the validity of the other cue and the cue intercorrelation. Darlington [1968] gives a discussion of this and other measures of cue importance and points out that such measures are inappropriate when the independent variables are correlated.

For the special case of uncorrelated cues, the situation is much clearer. In this case, it is well known that the squared multiple correlation coefficient equals the sum of the squared independent-dependent variable correlations. Thus,

$$R_e^2 = r_{1e}^2 + r_{2e}^2 + \ldots + r_{ke}^2$$

and

$$R_s^2 = r_{1s}^2 + r_{2s}^2 + \ldots + r_{ks}^2 .$$

Also, for the two cue case mentioned above, the regression coefficients reduce to:

$$\beta_1 = r_{1e} \frac{\sigma_{y_e}}{\sigma_{x_1}} .$$

$$\beta_2 = r_{2e} \frac{\sigma_{y_e}}{\sigma_{x_2}}$$

Thus, when the cues are uncorrelated, either the regression coefficients or the cue utilization coefficients, r_{is}, may be employed as measures of the extent to which a particular cue is utilizes. Because of their direct and obvious relationship to the multiple correlation coefficient, the cue utilizations are used here.

Cognitive Strain and Cue Utilization

In a previous section, the concepts of "cognitive strain" and "cognitive economy" were mentioned. Roughly, cognitive strain refers to the stress or hardships placed on the information processing mechanism of the individual due to the form, structure or amount of information provided. Cognitive economy is achieved by the use of simplifying strategies which reduce cognitive strain. As mentioned previously, Bruner, Goodnow and Austin [1956] present a lucid account of the relevance of these concepts in human information processing at the conceptual level.

In the context of multiple cue probability learning or inference tasks, cognitive economy may be achieved through the use of cue weighting policies which facilitate information processing. A good example of the nature of cognitive strain and a strategy which results is seen in the work of Slovic and McPhillamy [1974]. In describing the task, they say,

> Consider a task in which a judge must compare two students with respect to the criterion of potential college grade point average (GPA). He is given each student's score on two cue dimensions on which to base his judgments. One dimension is common to both students but the other dimension for each student is unique.

The difficulty is that different types of judgments are required for each dimension if one is to judge which student will have the higher GPA and by how much. Along the common dimension a direct comparison or relative judgment can be made. With respect to the unique dimension, an absolute judgment is required which is, of course, much more difficult. Slovic and McPhillamy found that the strategy consistently adopted was to weight the common dimension more heavily than the unique dimensions. Presumably, this policy simplified the task significantly as it was used even when explicit instructions were given to weight each dimension equally.

In the example just described, the source of the cognitive strain was the structure of the information presented and the nature of the judgment required. The problem under consideration in the present work is the strain induced by the amount of information presented and strategies which may reduce the strain. Miller [1960] has classified seven possible response strategies for dealing with information overload conditions. Among these categories are omission, simply ignoring some of the information, and filtering, a selective scheme for ignoring some information. Thus, a judge given cues A and B may employ omission by simply disregarding B or may filter by saying that B is important only for certain levels of A.

In the basic multiple cue probability learning task using cues which are pure numbers without labels, the filtering approach seems unlikely unless it is clearly suggested by the results of earlier trials. The subject might adopt such a strategy in order to improve performance if this is indicated by the results of previous trials, but the adoption of such a policy solely to reduce an overload condition seems doubtful. A much more plausible strategy is simply to ignore one or more of the cues so that attention may be focused on a smaller cue set. In this case, the subject would presumably attempt to find the subset which best enables him to predict the criterion if one exists. When the cues are uncorrelated, this amounts to choosing the subset of a given size with the highest validities. If feedback about the results of previous judgments is given, the subject would tend to search for a subset which best enables him to perform. Thus, it is not unlikely that the judgment policy would be rather unstable, at least in early trials. If the cues are of unequal validity or if no given subset of cues is more predictive than other subsets of the same size, a search strategy would prove fruitless and one would expect the subject either to abandon the omission strategy, to arbitrarily choose a subset of cues utilized or to utilize different cues from trial to trial on a more or less random basis.

Cue Utilization in Experiment Two

Experiment two contained uncorrelated cues so that it is possible to analyze the results of that experiment with respect to cue utilization. Appendix G contains the cue utilization coefficients for each subject and each condition of that experiment where the data were analyzed in one block of 150 trials. Also presented there are the values of R_s^2 for each subject which, as would be expected, are somewhat less than the average values reported in the previous chapter. This is partially due to the normal shrinkage that would be expected as the number of trials is increased and partially due to the fact that a single model based on all trials would fail to reflect policy changes which might occur as the experiment progresses. Thus, a subject who, dissatisfied with his performance in a subset of trials, adjusts his strategy would not be fit well by a single model.

Since the cues in this experiment are of approximately equal validity, the experiment is not particularly diagnostic with respect to cue utilization. In other words, there is no reason for the subject to utilize a particular cue or subset of cues to the exclusion of the others. Even if he should choose to do so, it is unlikely that he would persist in such a strategy since it would prove rather fruitless. A searching strategy which utilizes a small number of cues for a series of trials and then adopts a different subset would result in uniformly low cue utilizations.

Under these conditions, it seems unlikely that the cues would be utilized differentially when analyzed over the entire experiment. For this reason, one is interested in testing the hypothesis that all of the cue utilizations are equal, i.e., $r_{1s} = r_{2s} = \ldots = r_{Ks}$. A failure to reject this hypothesis would indicate a uniform pattern of cue utilization across cues.

For the data in experiment two, each of the cue utilizations is transformed to a z value so that $r_{is} = z_i$. It can be shown (see Graybill [1961], p. 210, Theorem 10.19) that the quantity

$$W = \sum_{i=1}^{K} (n_i - 3)(z_i - \bar{z})^2$$

follows approximately a chi-square distribution with K - 1 degrees of freedom where n_i is the sample size for the *ith* correlation coefficient $\bar{z} = \sum_{i=1}^{K} (n_i-3)z_i / \sum_{i=1}^{K} (n_i-3)$, and the populations are bivariate normal. The results of these tests are summarized in Table 14.

It is quite clear from the results in Table 14 that there is no consistent pattern of unequal cue utilization. In fact the three significant differences observed differ little from the number which would be expected from a series of 30 such tests. One is forced to conclude that there is no evidence to indicate that subjects resort to the utilization of a small number of cues in order to reduce the amount of information processing required. If such behavior does in fact occur, the pattern is not maintained consistently over the experiment and thus the behavior cannot be observed by the rather macroscopic methods used here.

Table 14. Tests for Equal Cue Utilization in Experiment Two

	$R_e^2 = 0.50$		$R_e^2 = 0.80$	
2 Cues	W = 0.122	NS	W = 1.504	NS
	W = 1.030	NS	W = 0.436	NS
	W = 5.028	p<.025	W = 1.298	NS
	W = 1.475	NS	W = 0.022	NS
	W = 0.315	NS	W = 3.909	p<.05
5 Cues	W = 7.979	NS	W = 3.494	NS
	W = 8.064	NS	W = 7.839	NS
	W = 7.647	NS	W = 3.163	NS
	W = 4.868	NS	W = 3.977	NS
	W = 6.497	NS	W = 11.748	p<.025
8 Cues	W = 4.703	NS	W = 9.294	NS
	W = 6.983	NS	W = 3.914	NS
	W = 8.469	NS	W = 12.588	NS
	W = 9.032	NS	W = 3.304	NS
	W = 3.434	NS	W = 11.703	NS

It must be emphasized that these conclusions refer only to the case of equal validity cues as employed in this experiment. Thus, there is really no reason for the subject to favor one cue over another. It is quite likely that when the cue validities and/or weights are unequal, some such strategy is used and, perhaps, profitably.

Introduction to Experiment Three

The previous section analyzed the cue utilization data from experiment two. In that case, the number of cues was increased from two to five to eight at each of two levels of R_e^2. It is, of course, clear that this necessitates a decrease in cue validities as the number of cues increases. Thus, any possible effect due to cue validity is confounded with a number of cues effect.

In order to eliminate the confounding present in the cue utilization data, experiment three was designed to maintain cue validities constant as the number of uncorrelated cues is increased. It is immediately apparent that this leads to an increase in R_e^2 as a function of the number of cues. However, viewed in conjunction with the data from experiment two, this study should lead to a rather clear picture of the effect of increasing the number of equally valid cues on cue utilization.

The experiment was designed for cue validities of 0.33. This leads to R_e^2 values of 0.22, 0.56 and 0.89, respectively, for the two, five and eight cue cases. The theoretical and empirical of the task parameters are given in Appendix H. The theoretical means for each cue and for Y_e were 100 and the empirical values were within acceptable limits.

Procedure for Experiment Three

The subjects were 18 students taking an undergraduate course in industrial engineering. All were male and all were volunteers receiving neither pay nor course credit for participating. Six subjects were assigned to each of the two, five and eight cue problems. The instructions given to the subjects and the procedures followed were the same as for experiments one and two.

Analysis

It was hypothesized that r_a would be an increasing function of the number of cues since R_e increased. The performance measures plotted in Figure 8 suggest that this is, in fact, the case and the analysis of variance on r_a summarized in Table 15 shows that the effect is statistically significant.

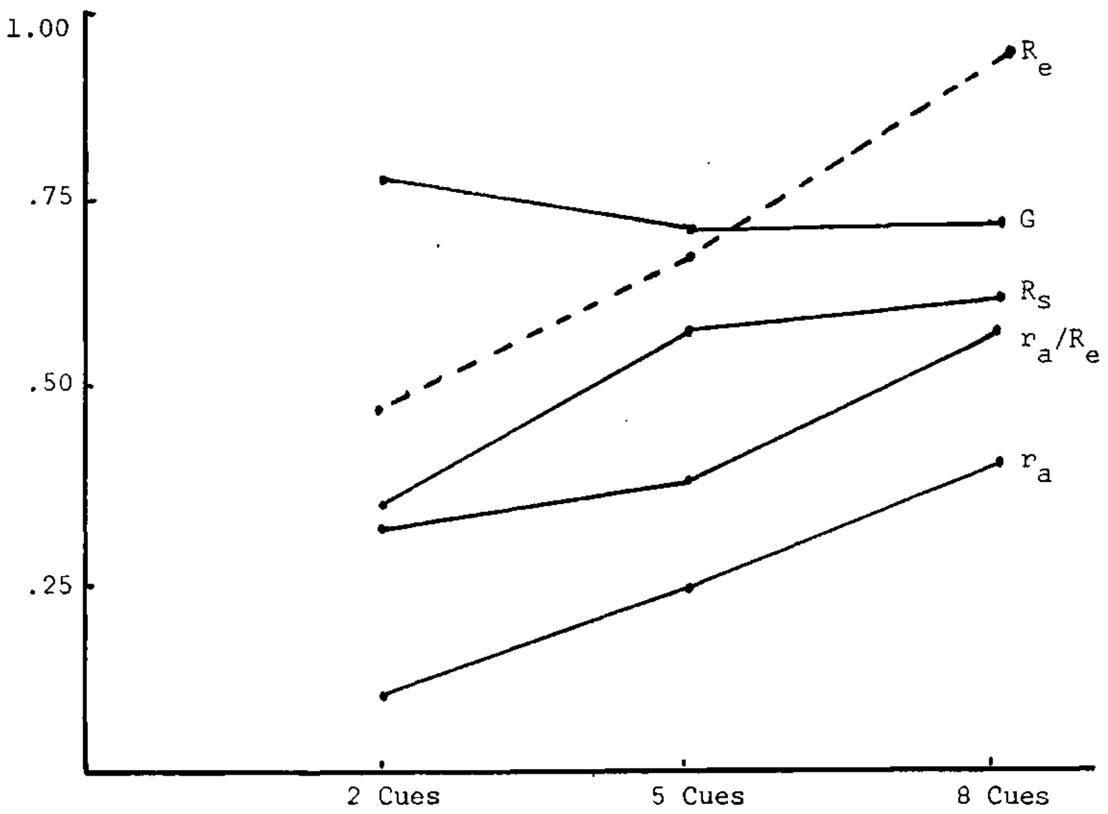


Figure 8. Performance Measures for Experiment Three

Table 15. ANOVA on r_a (Experiment Three)

Source of Variation	SS	df	MS	F
Between Subjects	2.636	17		
A (Number of Cues)	1.415	2	.708	8.698 ($p<.005$)
Subjects Within Groups	1.221	15	.081	
Within Subjects	1.397	36		
B (Blocks of Trials)	.124	2	.062	1.605
AB	.119	4	.030	<1
B x Subjects Within Groups	1.155	30	.038	

Tables 16 and 17 summarize the analyses of variance for G and R_s, respectively. It is noted that there are no significant effects in the analyses. A full summary of response data for experiment three is given in Appendix I.

Table 16. ANOVA on G (Experiment Three)

Source of Variation	SS	df	MS	F
Between Subjects	17.115	17		
A (Number of Cues)	.206	2	.103	<1
Subjects Within Groups	16.909	15	1.127	
Within Subjects	9.146	36		
B (Blocks of Trials)	.030	2	.015	<1
AB	2.354	4	.589	2.611
B x Subjects Within Groups	6.762	30	.225	

Table 17. ANOVA on R_s (Experiment Three)

Source of Variation	SS	df	MS	F
Between Subjects	9.206	17		
A (Number of Cues)	1.840	2	.920	1.874
Subjects Within Groups	7.365	15	.491	
Within Subjects	3.463	36		
B (Blocks of Trials)	.206	2	.103	1.094
AB	.427	4	.107	1.130
B x Subjects Within Groups	2.830	30	.094	

Having established that achievement does increase with the number of cues in this experiment, it is interesting to examine relative achievement (r_a/R_e). To accomplish this, a Kruskal-Wallis non-parametric one-way analysis of variance was conducted on relative achievement which was computed over 150 trials tested as a single block. Table 18 shows the results of this analysis which indicate that relative achievement did not differ significantly as a function of the number of cues and, hence, as a function of R_e.

The analysis of cue utilization in experiment three exactly parallels the analysis given previously for experiment two. Cue utilization coefficients were computed treating all 150 trials as a single block. The data are given in Appendix J. Table 19 summarizes the tests for equal cue utilization.

Table 18. Kruskal-Wallis Nonparametric One-Way ANOVA on Relative Achievement (Experiment Three)

2 Cues	5 Cues	8 Cues
-.009627 (1)	.096385 (4)	.389267 (10)
.452898 (14)	.015607 (2)	.383966 (9)
.026776 (3)	.236955 (7)	.569452 (17)
.430087 (12)	.173507 (6)	.491099 (15)
.168140 (5)	.753265 (18)	.370436 (8)
.449056 (13)	.408965 (11)	.532409 (16)
$R_1 = 48$	$R_2 = 48$	$R_3 = 76$
	H = 2.846 (NS)	

Table 19. Test for Equal Cue Utilization in Experiment Three

2 Cues	5 Cues	8 Cues
W = 0.383 NS	W = 18.040 (p<.001)	W = 14.215 (p<.05)
W = 6.695 (p<.01)	W = 4.240 NS	W = 8.055 NS
W = 0.108 NS	W = 12.770 (p<.025)	W = 21.228 (p<.005)
W = 8.559 (p<.025)	W = 0.950 NS	W = 19.681 (p<.01)
W = 0.021 NS	W = 3.155 NS	W = 15.154 (p<.05)
W = 42.735 (p<.005)	W = 3.367 NS	W = 14.870 (p<.05)

Out of 18 individual tests, ten subjects displayed significantly unequal cue utilization. In fact, of six subjects in the eight cue condition, five were seen to exhibit cue utilization patterns which differed significantly from equality. The overall results here differ from those for experiment two. It appears that there is no general trend towards the utilization of a subset of cues as the number of cues increases, but rather, the phenomenon is specific to individuals. In other words, the adoption of an unequal utilization policy when confronted with equal validity cues is a rather personal response to an information overload situation and is not highly dependent on the number of cues or the cue validities.

Summary

The analysis of the cue utilization coefficients in experiment two revealed little or no tendency to differentially utilize uncorrelated cues which are of equal validity. However, experiment three revealed a higher degree of reliance on such a policy although the use of the strategy was not related in any clear fashion to the number of cues available. This leads to the tentative conclusion that this type of response pattern is a function of the individual and not of the task parameters.

In tasks such as those used here, there is no a priori reason to believe that subjects would utilize any particular subset of cues to the exclusion of the others. Even if such a policy is adopted, it may not be maintained consistently over a large number of trials since no subset of a given size is more predictive than any other subset of the same size. This may lead to a searching strategy which would obscure any differences in cue utilization. One would expect that differential cue validities or certain verbal contexts which suggest the importance of certain cues would lead to differential utilization.

CHAPTER V

HIERARCHICAL INFERENCE TASKS

The Structure of Inference Tasks

All of the tasks considered thus far have involved criterion variables and cues which are related by simple functional relationships such as $X_i = Y_e + \epsilon_i$ or $Y_e = X + \ldots + X_k + \epsilon$. This structure implies that each cue is a variable directly related to the criterion of interest. In a working paper, Connolly [1974] has argued that several distinct models are available which may be appropriate for different situations. Consider, for example, the case of a psychologist who is interested in estimating the IQ of individuals. Several different instruments may be available, each of which measures IQ but with some error. An appropriate model of the process may thus be $X_i = Y_e + \epsilon_i$ where X_i is the observed score on the *ith* test, Y_e is the true IQ of an individual and ϵ_i is a random error component specific to the *ith* test. The variance of the score on the *ith* test, then, is assumed to be composed of true score variance and independent error variance. Since the true score variance is common to each measurement, scores on different tests will be correlated. Connolly refers to this as a single underlying variable (SUV) task.

In contrast, consider the marketing manager who wishes to estimate total sales revenue for a large corporation which has two autonomous divisions. In this case, total sales revenue would be estimated by adding the sales estimates for each division. The appropriate model would be $Y_e = X_1 + X_2 + \epsilon$ where Y_e is total sales revenue, X_1 and X_2 are the sales revenues for divisions one and two, respectively, and ϵ is a random error term. In this case, the independent variables may or may not be correlated. The fundamental distinction between this case, which Connolly refers to as the single resultant variable (SRV) model, and the SUV model discussed above is that for SUV tasks each independent variable is in fact a cue to the dependent variable whereas for SRV tasks each independent variable is a component of the dependent variable. Thus, the marketing manager can say little about total sales revenue unless he has an estimate of each component of revenue. The reader is referred to Connolly [1974] for an extended discussion of this distinction.

The foregoing discussion of SUV and SRV tasks is basically concerned with the direction of causality which, of course, will not affect the construction of a regression model but which may be a significant behavioral variable. Either cues or components may be elements of a more elaborate task structure than that considered thus far. In particular, these elements may be arranged in a hierarchical fashion which requires inference to be made on several levels. Figure 9 illustrates such a situation with Y_e being the ultimate criterion variable which is to be estimated, the X_i being the observable independent variables and the $A_i, B_{ij}, \ldots$ being intermediate constructs.

Tasks of a similar structure have been examined under the Bayesian paradigm. The general finding has been that when inferences are cascaded, they tend to be excessive compared to the normative model based on Bayes' Theorem. This result is in contrast to usual conservatism finding of single-stage Bayesian research. The reader is referred to Schum, Du Charme and Pitts [1973] and Gettys, Kelly and Peterson [1973] for examples of this methodology.

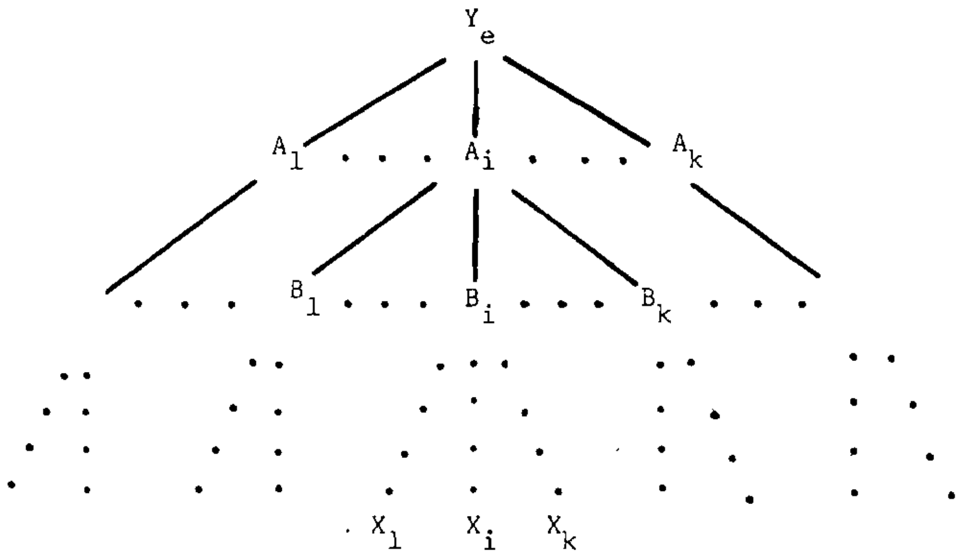


Figure 9. Structure of an Hierarchical Inference Task

A Two-Level Inference Model

As a particular two-level model consider a task involving two components of Y_e and two cues to each component. This task may be illustrated by a hypothetical example used by Frey [1974]. This example involved the estimation of suitability of a job applicant. The job was assumed to be composed of a business skill component and a personality or behavioral component. Cues to the business skill component were hypothetical ratings of the applicant by two experienced businessmen while cues to the behavioral component were ratings by psychologists.

The task structure was that overall efficiency was the product of a "true" business rating and a "true" psychological rating. The observed ratings were formed from the "true" ratings plus an independent error term. This structure is schematically illustrated in Figure 10.

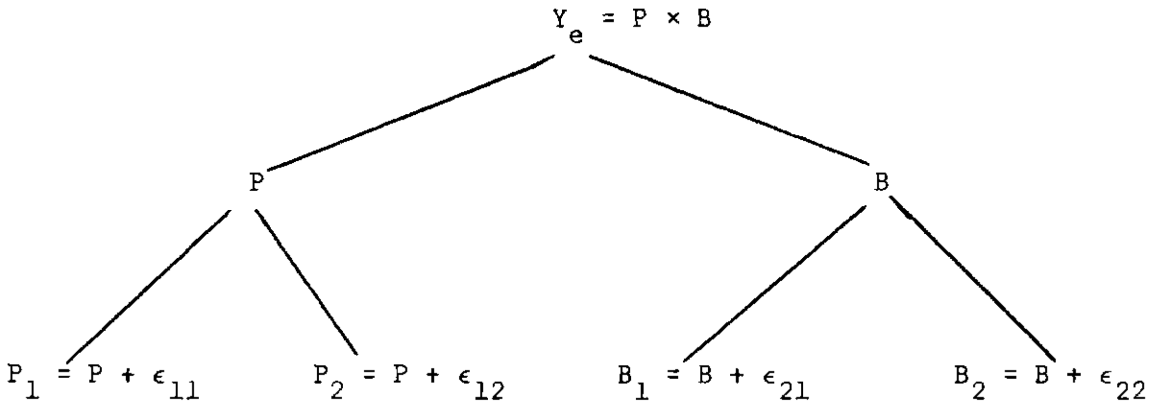


Figure 10. Task Used by Frey [1974]

Frey used three groups each of which received different degrees of insight into the task structure. The subjects in group one were told to utilize the four ratings in order to estimate overall efficiency. The second group was told that two ratings were by psychologists and two by businessmen with the rating labelled appropriately. The final group was given this information and was told to give an overall business rating, and an overall psychological rating as well as an overall efficiency rating. The idea was to force the subjects to structure their inference process by at least noting the intermediate constructs. All groups received outcome feedback after each trial. The results indicated that insight may increase performance in such a task although the results seem to depend on several factors including the shape of the criterion variable distribution.

Introduction to Experiment Four

In order to explore the effects of an increased number of cues in a two-level hierarchical task such as that described above, a fourth and final experiment was designed. Two problems were used with structure similar to that used by Frey discussed above. In each case, the criterion variable was formed as the product of two intermediate variables. For problem one, there was a total of four cues, two relating to each intermediate variable while problem two involved eight cues with four for each intermediate variable.

The hypothesis was that increased insight into the task structure or increased emphasis on the intermediate variables would lead to higher achievement. This hypothesis is based on the supposition that subjects who are not informed of the presence of the intermediate variables will have greater difficulty in discovering the correct strategy than subjects who are told of the intermediate variables and those who are forced to actually estimate values of the intermediate variables are more likely to detect and use the optimal strategy. It was further hypothesized that achievement, or at least relative achievement, would be lower for the eight cue problem than for the four cue problem. This, of course, is based on the notion that larger cue sets will tend to impede the learning of correct strategies.

Task Development

The model for problem one in experiment four was

$$Y_e = C \cdot S$$

$$C_1 = C + \epsilon_{11}$$

$$C_2 = C + \epsilon_{12}$$

$$S_1 = S + \epsilon_{21}$$

$$S_2 = S + \epsilon_{22}$$

where

$$C \sim N\ (20,3^2)$$

$$S \sim N\ (20,3^2)$$

$$\epsilon_{ij} \sim N\ (0,5^2) \qquad i=1,2;\ j=1,2.$$

C, S and the ϵ_{ij} are mutually independent. It can be shown that if C and S are independent, $C \cdot S$ has mean $\mu_c\mu_s$ and variance $\sigma_s^2 + \sigma_c^2 + \mu_s^2\sigma_c^2 + \mu_c^2\sigma_s^2$. Thus, $C \cdot S$ has mean 400 and variance 7281. This model yields

$$R_{sa}^2 = R_c^2 = \frac{4(9)}{4(9) + 2(25)} = \frac{36}{86} = .418$$

and

$$R_{sa} = R_c = .647.$$

Problem two had exactly the same structure except that there were four cues for each intermediate component and each ϵ_{ij} had a variance of 50. This leaves the predictability of each intermediate variable, R_s^2 and R_c^2, unchanged. It should be noted that the second stage of the process which involves combining C and S multiplicatively to yield Y_e is deterministic for known values of C and S. In both problems, the predictability of C and of S remain the same even though the number of cues changes. This yields a direct equivalence between the four cue task and the eight cue task even though the values of R_e^2 differ. The empirical values of the cue intercorrelations, cue validities for C, S and Y_e and R_c^2, R_{sa}^2 and R_e^2 are given for each problem in Appendix K.

Procedure

Forty-two undergraduate students taking a course in industrial engineering were unpaid subjects for this experiment. Some of them had participated in one of the three previous experiments. They were divided into six groups of seven members each which corresponded to three levels of instructional insight for each of the two problems. The three groups for each problem were run simultaneously in different rooms and both problems were run on the same day.

For problem one, instruction set I told the subject that the task was to estimate the overall desirability of automobile passenger restraint systems on the basis of four ratings of aspects of the sytem labelled X1, X2, X3 and X4. They were given no further insight into the nature of the ratings or the relationships between the cues or the cues and the criterion. They were given five practice trials in order to get some idea of the scales used and the pace of the task.

Instruction set II informed that two of the given ratings, labelled S1 and S2, referred to the safety of the system while the other two, labelled C1 and C2, referred to comfort. The subjects were asked to rate the overall desirability of the system and were given five practice trials.

Instruction set III was identical to set II except that the subjects were required to give an overall safety rating, an overall comfort rating and an overall desirability rating. This group was also given five practice trials.

The groups who received problem two were given exactly analogous instructions except that eight cues were used instead of four. Thus, for instruction set I, the cues were labelled X1, X2, ..., X8 and for sets II and III, the labels were S1, S2, S3, S4, C1, C2, C3, C4. The exact instructions given to the problem one subjects differ only in the manner described above. Each subject received an instruction sheet which was also read aloud by the experimenter after which procedural questions were answered. A sample instruction set is given in Appendix L.

Each group had 80 trials each requiring approximately 30 seconds each. The total time for distribution of materials, reading of instructions and conducting the test required slightly less than one hour. All subjects were given trial outcome feedback which was recorded in a space provided on the answer sheets. The cues were projected onto a screen in the manner described for the previous experiments.

Results and Analysis

The results of the experiment are given in Appendix M which contains r_a, G and R_s for each subject in each group in two blocks of 40 trials. Mean values of these quantities averaged over subjects and blocks of trials are plotted in Figure 11.

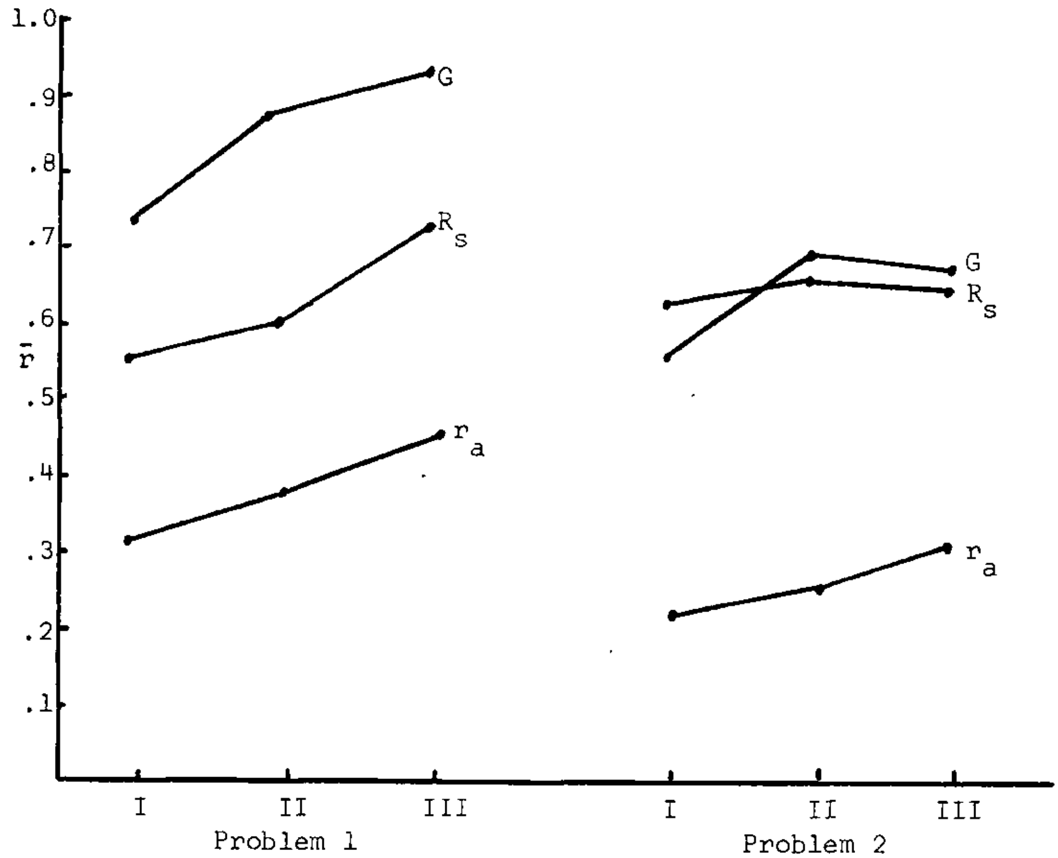


Figure 11. Performance in Experiment Four

The analysis of variance on r_a is summarized in Table 20. It will be noted that even though there appears to be a slight increasing trend with level of instruction, this effect does not reach statistical significance. In fact, the only significant effect is a number of cues by blocks of trials interaction which is illustrated in Figure 12.

Table 20. ANOVA on r_a (Experiment Four)

Source of Variation	SS	df	MS	F
Between Subjects	3.65	41		
A (Number of Cues)	.25	1	.25	2.90
B (Instructions)	.31	2	.15	1.78
AB	.02	2	.01	1
Subjects Within Groups	3.08	36	.09	
Within Subjects	1.14	42		
C (Blocks of Trials)	.01	1	.01	1
AC	.26	1	.26	12.66
BC	.04	2	.02	1
ABC	.09	2	.04	2.10
C x Subjects Within Groups	.75	36	.02	

Even though Figure 11 suggests that achievement in problem one (four cues) is higher than achievement in problem two (eight cues), the analysis of variance shows that the difference is not significant. The values of R_e, however, differ for the two problems and this suggests that relative achievement (r_a/R_e) may be higher for problem one than for problem two. To test this, the data were run in one block of 80 trials and r_a computed for each subject and R_e computed for each problem. These results are plotted in Figure 13 and are included in Appendix N. The distribution-free Wilcoxon rank sum test large sample approximation was used to test the one-sided hypothesis that relative achievement in problem one is equal to that in problem two (see Hollander and Wolfe

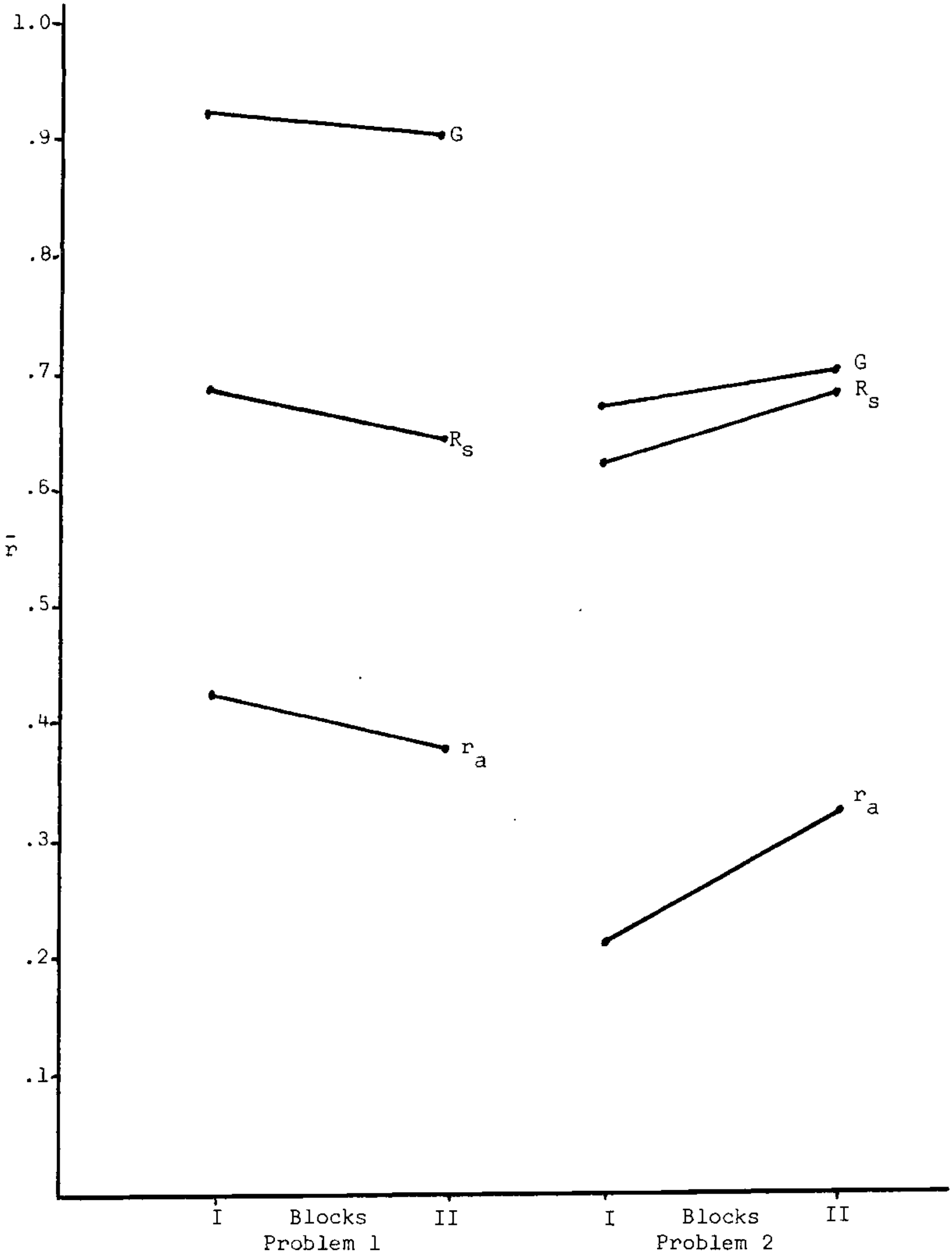


Figure 12. Performance in Experiment Four by Blocks of Trials

[1973], p.68). The hypothesis was rejected ($W^* = 2.50$, $p<.01$) indicating that relative achievement is greater in problem one.

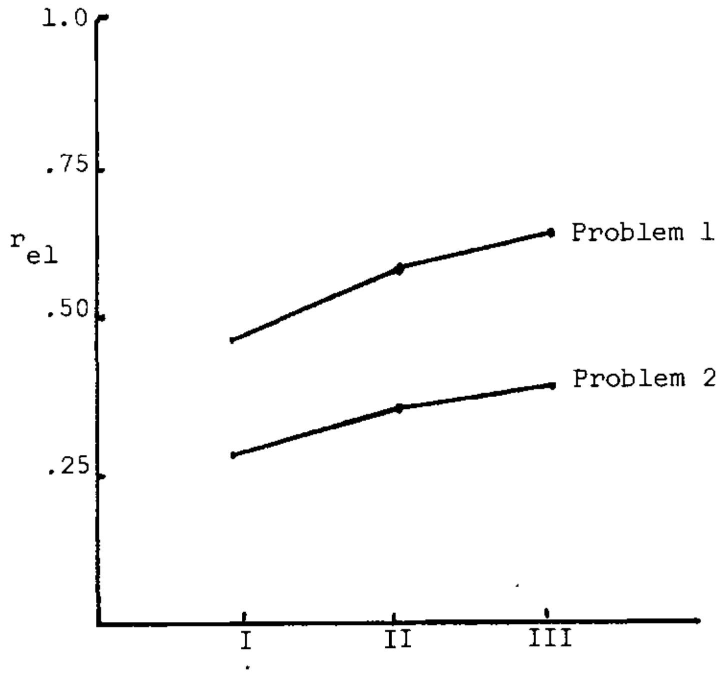


Figure 13. Relative Achievement (r_a/R_e) for Instruction Sets I, II and III of Experiment Four

The analyses for G and R_s are presented in Tables 21 and 22, respectively. The effects are plotted in Figure 11. It will be noted that G is significantly lower for problem two than for problem one and that the only significant effect for R_s is a problem by blocks of trials interaction. It must be remembered that G and R_s are based on first-order models, i.e., linear combinations of the cues. The optimal model of the environment, however, involves cross-product terms. Thus, G is only a measure of the extent to which the first-order model of the subject involves equal weights and R_s is a measure of the fit of this model. It appears that subjects' responses are fit reasonably well by this model.

Table 21. ANOVA on G (Experiment Four)

Source of Variation	SS	df	MS	F
Between Subjects	35.56	41		
A (Number of Cues)	10.93	1	10.93	18.75
B (Instructions)	2.53	2	1.26	2.17
AB	1.10	2	.55	1
Subjects Within Groups	20.99	36	.58	
Within Subjects	5.95	42		
C (Blocks of Trials)	.01	1	.01	1
AC	.23	1	.23	1.54
BC	,15	2	.08	1
ABC	.28	2	.14	1
C x Subjects Within Groups	5.27	36	.15	

Table 22. ANOVA on R_s (Experiment Four)

Source of Variation	SS	df	MS	F
Between Subjects	6.17	41		
A (Number of Cues)	.01	1	.01	1
B (Instructions)	.62	2	.31	2.10
AB	.23	2	.11	1
Subjects Within Groups	5.31	36	.15	
Within Subjects	1.84	42		
C (Blocks of Trials)	.01	1	.01	1
AC	.42	1	.42	11.41
BC	.01	2	.01	1
ABC	.08	2	.04	1.01
C x Subjects Within Groups	1.33	36	.04	

One additional observation can be made for the sake of completeness. Frey [1974] has pointed out that R_e is not the upper limit on achievement in this case since the optimal model does not involve a simple linear combination of the cues. For this reason, he defined R_e^* to be the correlation between Y_e and the product of the least-squares estimates of the two components of Y_e. Thus, if S and C are, respectively, the least-squares estimates of the safety component and the comfort component, R_e^* is the correlation between Y_e and Y_e^* where $Y_e^* = S \cdot C$. The values of R_{sa}^2, R_c^2 and R_e^{*2} are given in Appendix K.

Figure 14 shows relative achievement analogous to Figure 13 but With R_e^* used instead of R_e. Figure 13 is identical to Figure 14 except for a rescaling of the ordinate since R_e^* has a constant ratio to R_e.

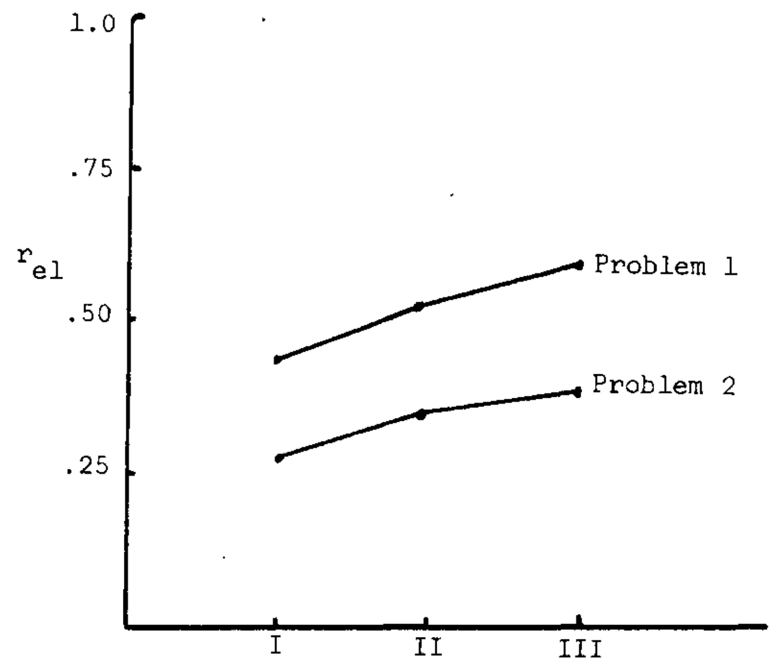


Figure 14. Relative Achievement Using R_e^* for Instruction Sets I, II and III of Experiment Four

Discussion

This experiment did not show a significant increase in achievement as a function of task insight. It is of interest to note a trend toward higher achievement (see Figure 11) at each number of cues. That this trend is not statistically significant suggests that there is a rather large error variance, i.e., large individual differences. This is supported by reference to the relative achievement (r_a/R_e) data given in Appendix N. For problem one, relative achievement ranges from 0.062 to 0.921 with a median of 0.627; the range for problem two is -0.092 to 0.749 with a median of 0.351. It should be noted that these figures are based on 80 trials and thus should be quite indicative of individual performance.

While achievement did not differ for the two problems, there was a significant difference in relative achievement with a higher level for problem one (four cues) than for problem two (eight cues). It would probably be an error to attach too much significance to this result. Basically, subjects seem more efficient with smaller numbers of cues but the overall task predictability is so low (the best linear combinations of the cues explain about 40 per cent of the variance in problem one and about 55 per cent in problem two) that subjects seem to have great difficulty in detecting the correct strategy. Informal post-experiment discussions with the subjects tend to support this finding.

The difficulty of the problems used in this experiment probably explains the differences in the findings reported here and those of Frey [1974]. Frey did obtain a significant insight effect for a task with $R_e = 0.96$. A task with this level of predictability provides the subject with ample opportunity to detect the optimal strategy. The sensitivity of hierarchical inference tasks to task parameters is shown by another study conducted by Frey using a different form of Y_e distribution which yielded ambiguous results with respect to insight.

One is left with the interesting conjecture that insight may be more important in relatively simple tasks than in relatively difficult tasks or, at least, that insight given must receive some support from increased success in the problem. When the task structure contains a good deal of randomness, even subjects who hit upon the best policy may not attain a level of performance which convinces them that they are doing as well as is possible. This possibility is suggestive of the following passage from Edwards [1971]:

> While we complain of human irrationality, we would probably also (if not too bigoted) admit that the following question and answer contain a powerful, rather accurate, and often operational theory of human behavior:
>
> Q: What is he doing?
> A: He's doing the best he can.

CHAPTER VI

CONCLUSIONS FROM THE RESEARCH AND AREAS FOR FURTHER STUDY

A Restatement of Purpose and Scope

The general area of interest addressed by the studies presented here concerns the effects of increased amounts of information on human judgment or decision making performance. Motivation for an interest in the area arises on both a theoretical or scientific level and on a practical level. One is interested in the ultimate limit on the human capacity to assimilate and process information. On the other hand, a more thorough understanding of man's reaction to information overload situations could lead to more efficient designs for decision making systems such as management information systems. In particular, it could lead to revised philosophies for allocating decision making tasks to man and to machine.

Within the general area of reaction to information overload situations, the studies presented here are concerned with the effects on performance of varying the number of cues available to the decision maker in a multiple cue inference task. The studies were designed and analyzed within the conceptual framework afforded by the lens model and were conducted in laboratory rather than field settings. As is always the case in behavioral research, this creates some difficulty in extending the findings to situations commonly encountered in the "real-world."

However, there is the advantage that the findings are the result of carefully controlled experimental conditions.

The specific problems that this research set out to investigate were:

1. With task predictability held constant, how does performance change as a function of the number of cues available to the decision maker?

2. Does the pattern in (1) depend on total task predictability or degree of cue redundancy?

3. Is there evidence that as the number of cues increases, subjects tend to rely on only a few cues rather than the entire set?

4. If the task has a complex, hierarchical structure, does giving the subject greater insight into this structure lead to increased performance and does the effect of insight depend on the number of cues?

In each case, only equally valid cues were used.

Conclusions from the Research

Taken jointly, experiments one and two address the first two research problems mentioned above. With correlated cues, achievement (r_a) followed a non-monotonic pattern as a function of the number of cues for a high predictability task ($R_e^2 = 0.80$), but was not significantly affected by the number of cues for a low predictability task ($R_e^2 = 0.50$). When uncorrelated cues were employed, there was a general and statistically significant trend towards lower achievement with large cue sets. In this case, the shape of the function relating achievement to the number of cues was essentially the same at each level of task predictability. It is of interest to note that when correlated cues were employed, achievement and relative achievement were higher for the high predictability task than for the low predictability case. However, with uncorrelated cues, there was no significant task predictability effect on r_a. With regard to two-cue problems, Naylor and Schenck [1968] found that achievement increased as the degree of cue intercorrelation increased and that the facilitating effect of increased cue redundancy was greater for larger values of R_e. Analyses of experiments one and two presented here show achievement to be significantly higher for correlated than for uncorrelated cues and the difference is much more dramatic for $R_e^2 = 0.80$ than for $R_e^2 = 0.50$. Thus, the Naylor and Schenck finding is extended to five-cue and eight-cue problems.

Regarding cue utilization, there was apparently no trend towards more use of a subset of cues as the number of cues increased. It had been hypothesized that this type of policy might result from an information overload situation. Since, however, all cues were of equal validity, there was little reason for the subject to utilize any particular subset of cues to the exclusion of the others. The general finding was that the adoption of such a strategy seems to be dependent on the particular individual involved and not on the amount of information or the task parameters.

The final study examined the effects of no insight, suggested insight and forced insight on performance in an hierarchical inference task. The insight related to the structure of the task involved. The findings suggested that although there was a tendency for increased insight to lead to higher achievement, the effect did not reach statistical significance. Achievement was higher for the four-cue task than for the eight-cue task, but relative achievement did not differ.

It should be pointed out that all of the studies above related to the case in which the amount of information provided to the decision maker varied independently of the amount of data provided. That is, if the number of cues is increased while the total task predictability is held constant, the subject receives more data but not more information. The amount of overall information, as related to an upper bound on achievement, is manipulated by varying the task predictability. A central focus has been on the ability of the subject to process more data with the total amount of information held constant. In this sense, one is concerned with human data processing in addition to human information processing.

An Outline for Further Research

Studies of the human judgment/inference process may vary along at least two dimensions--a task dimension and a subject dimension. Tasks may run from rather artificial laboratory studies involving pure members to complex field studies employing "real-world" problems. Subjects may vary from being naive with respect to the particular problem to being acknowledged experts. Connolly and Dorris [1974] explore these dimensions more fully. The suggestions for future research to be outlined here will concentrate primarily on task dimensions.

The studies presented here have to a large extent exhausted the set of pure laboratory studies employing equal validity cues and concentrating on varying the number of cues available to the decision maker. It would, however, be of interest to examine more fully the relationship between cue intercorrelations and achievement as the number of cues is increased. The results presented here and the study by Nystedt and Magnusson [1972] indicate that cue redundancy plays a significant role. This suggests studies that systematically vary the degree of cue intercorrelation in a manner similar to Naylor and Schenck [1968].

The more interesting and more general problem relates to the use of problems with differential cue validities. For example, within the realm of laboratory studies, it would be of interest to replicate experiments one and two of this report using problems with differential cue validities. Such tasks are more likely to induce subjects to weight the cues unequally. It might be hypothesized that these tasks would lead to higher achievement than would tasks involving equally valid cues since differential validities would lead to the use of a small subset of the available cues rather than the entire set. Proper weighting of this subset might well be more efficient than improper utilization of the entire cue set.

In conjunction with work along these lines, one might look at subjective self-policy assessment and insight into policy as a function of the number of cues (see Slovic and Lichtenstein [1971] for a review of work in this area). It may well be that judges are able to appropriately weight a small number of cues but do less well as the number increases.

The regression approach and the subjective estimate approach to policy capturing are rather indirect. A particularly promising alternative is to study information search and purchase patterns. If the judge is motivated to perform well and must purchase information upon which to base his judgments, the record of his purchases would provide a direct estimate of his evaluation of the cues. In addition, purchase behavior over a series of trials would indicate his information search strategy. It is conjectured that as the number of cues is increased, more elaborate search strategies would be required. This approach has the additional advantage of reducing a large number of cues (those available) to a smaller number (those purchased). Initial research in this area is in progress by Connolly and Woo [1974].

The studies reported here all involved an observable criterion variable, the value of which the subject was to estimate and about which the subject received feedback. Some other studies (for example Einhorn [1971], have looked at tasks which do not involve an observable criterion value and which thus do not involve feedback to the judge or decision maker. Typically, these tasks involve preferences or utilities rather than inference. An example would involve the estimation of job desirability on the basis of pay, location and rank. Additionally, these preference tasks are often self-paced whereas inference tasks are often externally-paced. There are certain obvious structural differences such as the fact that achievement, cue validity, and task predictability are not defined for preference tasks. The behavioral distinctions, however, are not so clear. Perhaps the primary distinction is that research on preference tasks is basically concerned with policy capturing and the fit of various types of models while research on inference tasks tends to focus on achievement and learning. As an example of the differences, one would expect a regression model to provide a better fit for a preference task than for an inference task since, in the former case, the judge receives no feedback which might tend to induce revisions in his policy. The basic premise in an inference task, on the other hand, is that learning will occur which will lead to revised policies and hence to poorer fit of a given model. These distinctions need to be more fully understood since some of the more important work on the effects of increased amounts of information have involved preference tasks.

A particularly promising area for future research is related to an ambiguous finding reported here. It seems very plausible that for hierarchical tasks, structural insight would lead to higher performance and an increased ability to handle large cue sets (see Frey [1974]). In experiment four of this research, results were obtained which were consistently in the correct direction but failed to reach statistical significance. Evidently, this effect is extremely sensitive to task parameters. The required studies would involve a wide range of tasks and a more comprehensive look at the types of structural insight which might be used.

All of the suggested studies outlined above could be conducted in laboratory settings. Quite possibly, however, the ultimate payoff for judgment research will be in the study of true experts functioning in their natural environments. It would be of great interest to discover if experimental findings such as those reported here can be extended to "real-world" situations. One frequently hears reference made to the "information explosion" and the "fact" that business executives, for example, are inundated by data. An empirical verification of such an information saturation point would be very interesting. Perhaps the basic question is: "Can it be established that a true expert, working in his normal environment and at his own pace, will make poorer decisions or judgments with large amounts of information than with smaller amounts?" This finding would have enormous implications for such diverse areas as medical diagnosis, weather forecasting and the design of management information systems.

Some Methodological Considerations

It has become increasingly clear that there are some rather severe difficulties associated with the correlation-regression paradigm for the analysis of human information processing. With respect to the lens model and associated statistics, those which are based on regression models, R_s and G, are particularly suspect. For example, each of these measures considers a single model to be appropriate for a given block of trials. If, however, learning occurs, one would expect corresponding changes in the model. Of course, by reducing the number of trials in the block upon which the regression model is to be based, one may minimize the learning problem but only at the expense of reducing the ratio of number of observations to parameters to be estimated. This may result in unstable regression coefficients.

The measure G has specific difficulties related to it. G, the matching index, is a measure of the extent to which the regression coefficients of the paramorphic model of the judge are proportional to the optimal coefficients of the model of the environment. Alternatively, it is the correlation between the predicted values of the paramorphic and the predicted values of the optimal environmental model. Dawes and Corrigan [1974] have suggested that G is not very sensitive since, for the range of problems typically of interest, even sub-optimal models will correlate highly with the optimal model. They report a variety of simulation studies using unit weights and even randomly chosen weights which support their conclusion that as long as each independent variable has a conditionally monotonic relationship to the dependent variable, almost any model will correlate highly with the optimal one. The significance of this finding is that, rather generally, a model of the judge will correlate highly with the environmental model and, thus, G will be relatively large.

The problems mentioned above are common to all studies employing this methodology. There are, however, difficulties specific to studies using large cue sets or cue sets of varying sizes. The most obvious difficulty relates to the somewhat delicate balance between keeping subjects motivated (or at least interested) and obtaining enough observations to construct a reasonable model. The problem is even more apparent if one wishes to subdivide the total set of observations into several blocks of trials.

Perhaps the most serious problem is that of comparing regression models based on different numbers of independent variables. For example, if conditions using two, five and eight cues are analyzed by means of an

ANOVA on G, one is actually comparing quantities based on two, five and eight independent variables. It is not at all clear how deviations from optimal weighting are compared for models of different sizes.

It is not readily apparent what type of methodology should be developed. It would certainly be desirable to employ such measures as eye movements as indicators of cue importance. Such measures are, however, task-specific and it remains to be seen if a new, more powerful conceptual framework can be developed to supplement the lens model. One specific area for which a new framework should be developed is the information search or purchase problem mentioned above. In this respect, it is likely that some of the economic decision models developed by operations research analysts would be appropriate. In a more general sense, the techniques of computer simulation could prove to be powerful tools.

In summary, the research reported here has provided baseline data for the study of the ability of subjects to process large cue sets. Further studies should investigate larger cue sets, cues of unequal validity and tasks which more closely approximate real-world decision problems. While the lens model paradigm was employed in those studies, there is substantial need for new methodological advances and models developed for particular classes of tasks.

APPENDIX A

DERIVATION OF LENS MODEL EQUATION

The purpose of this Appendix is to present the details of the derivation of the lens model equation and related quantities. The equation is presented and discussed in Chapter I. The development given here parallels that of Castellan [1973].

Let Y_e be the distal variable and Y_s be the subject's estimate of Y_e. Further, let

$$\underline{X}' = (X_1, X_2, \ldots, X_K)$$

be a vector of cues. Y_e, Y_s and $\underline{X}$ are random variables and no generality is lost by letting

$$E(Y_e) = E(Y_s) = 0$$

and

$$E(\underline{X}) = \underline{0}$$

where $\underline{0}$ is a vector of K zeroes. Clearly,

$$E(Y_e^2) = \sigma_e^2$$

$$E(Y_s^2) = \sigma_s^2$$

$$E(\underline{X}\underline{X}') = \underline{\Sigma}$$

$$E(Y_e\underline{X}) = \underline{\Sigma}\underline{e}$$

and

$$E(Y_s\underline{X}) = \underline{\Sigma}\underline{s}.$$

Assuming a model of the form

$$Y_e = \underline{\beta}_e'\underline{X} + \epsilon$$

with

$$E(\epsilon) = 0,$$

it follows that

$$L = E(\epsilon^2) = E[(Y_e - \underline{\beta}_e'\underline{X})^2]$$

$$= E[Y_e^2 - 2Y_e\underline{\beta}_e'\underline{X} + \underline{\beta}_e'\underline{X}\underline{X}'\underline{\beta}_e]$$

$$= \sigma_e^2 - 2\underline{\beta}_e'\underline{\Sigma}\underline{e} + \underline{\beta}_e'\underline{\Sigma}\underline{\beta}_e.$$

Differentiating and equating to zero,

$$\frac{\partial L}{\partial \underline{\beta}_e} = -2\underline{\Sigma e} + 2\underline{\Sigma \beta}_e = 0$$

$$\underline{\beta}_e = \underline{\Sigma}^{-1}\underline{\Sigma e}. \tag{1}$$

By the same argument,

$$\underline{\beta}_s = \underline{\Sigma}^{-1}\underline{\Sigma s}. \tag{2}$$

Let

$$Y_e = \hat{Y}_e + Z_e$$

and

$$Y_s = \hat{Y}_s + Z_s$$

where

$$\hat{Y}_e = \underline{\beta}_e'\underline{X}$$

and

$$\hat{Y}_s = \underline{\beta}_s'\underline{X}.$$

The correlation between Y_e and Y_s, designated by r_a, can be written:

$$r_a = \frac{\mathrm{cov}(\hat{Y}_e,\hat{Y}_s) + \mathrm{cov}(Z_e,Z_s)}{\sigma_e\sigma_s} \tag{3}$$

since

$$\mathrm{cov}(\hat{Y}_e, Z_e) = \mathrm{cov}(\hat{Y}_s, Z_s) = 0.$$

Now,

$$\mathrm{cov}(\hat{Y}_e, \hat{Y}_s) = r_{\hat{Y}_e, \hat{Y}_s} \left(\mathrm{Var}(\hat{Y}_e)\mathrm{var}(\hat{Y}_s)\right)^{1/2} \tag{4}$$

and

$$\mathrm{cov}(Z_e, Z_s) = r_{Z_e, Z_s} \left(\mathrm{Var}(Z_e)\mathrm{var}(Z_s)\right)^{1/2}. \tag{5}$$

By the definition of the multiple correlation coefficient,

$$\mathrm{Var}(\hat{Y}_e) = R_e^2 \sigma_e^2 \tag{6}$$

$$\mathrm{Var}(\hat{Y}_s) = R_s^2 \sigma_s^2 \tag{7}$$

$$\mathrm{Var}(Z_e) = (1-R_e^2)\sigma_e^2 \tag{8}$$

$$\mathrm{Var}(Z_s) = (1-R_s^2)\sigma_s^2. \tag{9}$$

Letting

$$G = r_{\hat{Y}_e, \hat{Y}_s} \tag{10}$$

and

$$C = r_{Z_e,Z_s} \tag{11}$$

and substituting (6) - (11) into (3), a new expression for r_a is:

$$r_a = \frac{GR_eR_s\sigma_e\sigma_s + C(1-R_e^2)^{1/2}\sigma_e(1-R_s^2)^{1/2}\sigma_s}{\sigma_e\sigma_s}$$

$$r_a = GR_eR_s + C(1-R_e^2)^{1/2}(1-R_s^2)^{1/2} \tag{12}$$

which is the lens model equation given in Chapter I.

APPENDIX B

THEORETICAL AND EMPIRICAL VALUES OF TASK PARAMETERS FOR EXPERIMENT ONE

Cues	R_e		Mean Cue Intercorrelation r_{X_i,X_j}		Mean Cue Validity r_{X_i,Y_e}	
	Theoretical	Empirical	Theoretical	Empirical	Theoretical	Empirical
2	0.70 0.90	0.72 0.90	0.33 0.67	0.33 0.68	0.58 0.82	0.59 0.83
5	0.70 0.90	0.73 0.91	0.17 0.45	0.19 0.47	0.41 0.67	0.40 0.69
8	0.70 0.90	0.74 0.92	0.11 0.35	0.10 0.32	0.33 0.59	0.36 0.61

APPENDIX C

INSTRUCTIONS FOR EXPERIMENT ONE

This study is designed to discover how people use information to make decisions. On the screen you will be given ___ numbers. You will also see the number of the trial which corresponds to the number on your answer sheet. On the basis of these data you will try to predict the value of another number. This estimate should be recorded next to the appropriate trial number under the column "Your Estimate." After your estimate has been recorded, I will show you the true value which you were attempting to estimate. This should be recorded under "True Value." We will then proceed to the next trial. You will not be able to predict perfectly, but you should do the best you can. The true values will be integers between zero and one thousand.

Any questions?

APPENDIX D

RESPONSE DATA FOR EXPERIMENT ONE

r_a

	$R_e^2 = 0.50$			$R_e^2 = 0.80$		
	2 Cues	5 Cues	8 Cues	2 Cues	5 Cues	8 Cues
S-1	.296787	.530046	.108024	.790380	.786172	.762432
S-2	.352313	.571132	.075173	.648198	.844912	.607485
S-3	.476586	.490952	.401970	.443454	.755519	.390599
S-4	.243778	.159327	.223270	.739407	.745621	.551472
S-5	.677785	.156761	.334053	.527760	.768204	.661037
S-1	.708604	.587753	.475063	.898972	.978655	.857137
S-2	.467701	.549061	.169972	.676497	.992199	.842685
S-3	.676217	.595802	.550214	.861588	.993487	.618808
S-4	.667827	.058702	.066562	.420422	.962278	.695145
S-5	.810524	.563441	.638162	.468077	.923638	.712597
S-1	.512033	.530755	.515204	.772746	.913875	.749144
S-2	.223304	.433417	.233307	.732464	.895456	.801412
S-3	.421886	.412859	.478861	.720232	.804362	.616052
S-4	.215738	.189830	.094098	.481915	.866907	.627525
S-4	.558821	.392641	.497737	.415486	.938653	.714799

		Means Based on Z's			
		Block 1	Block 2	Block 3	Means
$R_e^2 = 0.50$	2 cues	.425	.680	.396	.500
	5 cues	.341	.489	.397	.409
	8 cues	.232	.402	.371	.335
$R_e^2 = 0.80$	2 cues	.648	.721	.645	.671
	5 cues	.783	.830	.832	.815
	8 cues	.609	.760	.709	.693

R_s

		$R_e^2 = 0.50$			$R_e^2 = 0.80$		
		2 Cues	5 Cues	8 Cues	2 Cues	5 Cues	8 Cues
Block 1	S-1	0.476623	0.782299	0.333009	0.880725	0.913875	0.885000
	S-2	0.390998	0.765799	0.595549	0.645883	0.895456	0.808305
	S-3	0.791946	0.861725	0.714728	0.453541	0.804362	0.451543
	S-4	0.602977	0.243884	0.361874	0.763951	0.866907	0.804788
	S-5	0.960453	0.457327	0.589750	0.681649	0.938653	0.793980
Block 2	S-1	0.854828	0.805678	0.603470	0.942706	0.975597	0.956092
	S-2	0.607256	0.764749	0.515718	0.701039	0.907458	0.936509
	S-3	0.832234	0.839825	0.827083	0.917176	0.884776	0.767086
	S-4	0.771051	0.330994	0.555744	0.502493	0.915632	0.850298
	S-5	0.943287	0.798357	0.726696	0.548890	0.950160	0.819140
Block 3	S-1	0.855454	0.730064	0.745594	0.897150	0.919033	0.891721
	S-2	0.486249	0.728121	0.666576	0.869307	0.943584	0.940409
	S-3	0.905674	0.743665	0.741765	0.875484	0.858714	0.734586
	S-4	0.562571	0.376759	0.405283	0.733137	0.905427	0.813390
	S-5	0.960661	0.780467	0.661868	0.563191	0.869077	0.792938

		Means Based on Z's			
		Block 1	Block 2	Block 3	Means
$R_e \simeq 0.70$	2 cues	0.731	0.831	0.785	0.783
	5 cues	0.676	0.743	0.692	0.703
	8 cues	0.535	0.664	0.658	0.619
$R_e \simeq 0.90$	2 cues	0.714	0.791	0.816	0.773
	5 cues	0.892	0.935	0.904	0.911
	8 cues	0.778	0.887	0.853	0.839

G

		$R_e^2 = 0.50$			$R_e^2 = 0.80$		
		2 Cues	5 Cues	8 Cues	2 Cues	5 Cues	8 Cues
Block 1	S-1	0.886698	0.928301	0.516399	0.993909	0.978655	0.947580
	S-2	0.993027	0.969697	0.505739	0.983111	0.992199	0.902831
	S-3	0.976577	0.965537	0.802429	0.977420	0.993487	0.762933
	S-4	0.649107	0.418195	0.636162	0.997901	0.962278	0.731608
	S-5	0.995674	0.930852	0.632188	0.946371	0.923638	0.923985
Block 2	S-1	0.956120	0.895127	0.807054	0.998556	0.990007	0.973725
	S-2	0.990246	0.973867	0.525476	0.998756	0.987357	0.974189
	S-3	0.972316	0.965761	0.840358	0.999511	0.995654	0.837537
	S-4	0.992499	0.042193	0.428653	0.961383	0.997678	0.876996
	S-5	0.978382	0.946820	0.966915	0.995207	0.934036	0.975433
Block 3	S-1	0.999151	0.969237	0.899017	0.995822	0.982785	0.922303
	S-2	0.990133	0.925195	0.480401	0.992992	0.975550	0.922006
	S-3	0.905674	0.986460	0.778485	0.999711	0.985636	0.949586
	S-4	0.999910	0.530930	0.313972	0.874135	0.980820	0.829960
	S-5	0.993383	0.729161	0.877843	0.909730	0.980387	0.931618

		Means Based on Z's			
		Block 1	Block 2	Block 3	Means
$R_e \simeq 0.70$	2 cues	0.968	0.982	0.997	0.982
	5 cues	0.918	0.904	0.913	0.911
	8 cues	0.633	0.795	0.737	0.722
$R_e \simeq 0.90$	2 cues	0.988	0.997	0.990	0.992
	5 cues	0.980	0.990	0.981	0.984
	8 cues	0.879	0.949	0.918	0.915

APPENDIX E

THEORETICAL AND EMPIRICAL VALUES OF TASK PARAMETERS FOR EXPERIMENT TWO

(Empirical Values Above Diagonal--Theoretical Below)

CASE I: Two Cues, $R_e^2 \approx 0.50$

	X_1	X_2	Y_e
X_1		-.054	.503
X_2	.00		.466
Y_e	.50	.50	

Theoretical $R_e^2 = 0.500$

Empirical $R_e^2 = 0.497$

CASE II: Two Cues, $R_e^2 = 0.80$

	X_1	X_2	Y_e
X_1		.013	.658
X_2	.00		.619
Y_e	.62	.62	

Theoretical $R_e^2 = 0.800$

Empirical $R_e^2 = 0.805$

CASE III: Five Cues, $R_e^2 = 0.50$

	X_1	X_2	X_3	X_4	X_5	Y_e
X_1		.025	.079	.002	.013	.327
X_2	.00		.002	.046	-.031	.400
X_3	.00	.00		.012	-.047	.293
X_4	.00	.00	.00		.021	.335
X_5	.00	.00	.00	.00		.265
Y_e	.32	.32	.32	.32	.32	

Theoretical $R_e^2 = 0.500$

Empirical $R_e^2 = 0.508$

CASE IV: Five Cues, $R_e^2 = 0.80$

	X_1	X_2	X_3	X_4	X_5	Y_e
X_1		-.087	.064	.003	-.043	.347
X_2	.00		-.025	.002	.010	.384
X_3	.00	.00		.008	-.082	.397
X_4	.00	.00	.00		-.003	.414
X_5	.00	.00	.00	.00		.370
Y_e	.40	.40	.40	.40	.40	

Theoretical $R_e^2 = 0.80$

Empirical $R_e^2 = 0.781$

CASE V: Eight Cues, $R_e^2 = 0.50$

	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	Y_e
X_1		-.034	-.087	.015	.162	-.015	.125	.238	.286
X_2	.00		.042	-.015	.071	.034	-.099	.048	.243
X_3	.00	.00		.156	-.042	.032	.031	-.019	.269
X_4	.00	.00	.00		.084	.109	-.014	.052	.343
X_5	.00	.00	.00	.00		.089	.089	-.098	.316
X_6	.00	.00	.00	.00	.00		.064	.085	.313
X_7	.00	.00	.00	.00	.00	.00		.164	.357
X_8	.00	.00	.00	.00	.00	.00	.00		.419
Y_e	.25	.25	.25	.25	.25	.25	.25	.25	

Theoretical $R_e^2 = 0.500$

Empirical $R_e^2 = 0.626$

CASE VI: Eight Cues, $R_e^2 = 0.80$

	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	Y_e
X_1		.063	.047	-.027	-.155	.114	-.169	-.018	.207
X_2	.00		.017	.053	-.052	-.046	-.005	-.027	.324
X_3	.00	.00		.009	-.028	.011	-.021	-.086	.280
X_4	.00	.00	.00		.067	-.027	.093	-.028	.351
X_5	.00	.00	.00	.00		-.133	.057	.115	.272
X_6	.00	.00	.00	.00	.00		-.036	-.009	.312
X_7	.00	.00	.00	.00	.00	.00		-.052	.279
X_8	.00	.00	.00	.00	.00	.00	.00		.404
Y_e	.32	.32	.32	.32	.32	.32	.32	.32	

APPENDIX F

RESPONSE DATA FOR EXPERIMENT TWO

Experiment No. Two (Response Data) (r_a)

		$R_e^2 = 0.50$			$R_e^2 = 0.80$		
		2 Cues	5 Cues	8 Cues	2 Cues	5 Cues	8 Cues
Block 1	S-1	0.177313	.046420	-.125175	.450109	.306813	.018583
	S-2	.338226	-.024146	-.131542	.162146	.331860	-.016228
	S-3	.141783	.175535	-.065438	.216876	.411669	.281769
	S-4	.330786	-.161118	.236674	-.078999	.263972	.386897
	S-5	.630793	.499152	.021656	.237676	.177445	.065446
Block 2	S-1	.104634	.247080	.140570	.314807	.029750	-.143413
	S-2	.246259	.309059	.265010	.357359	.386924	.313121
	S-3	.251891	.205698	.055230	.551980	.233699	.077730
	S-4	.340454	.317847	.368833	.188599	.590140	.458289
	S-5	.073393	.390778	.226298	.450753	.363076	.157499
Block 3	S-1	.000000	.628190	.244080	.339852	-.011876	.197462
	S-2	.193944	.258318	-.225935	.366525	-.046690	.226216
	S-3	.195674	.184713	.351357	.522935	.522377	-.022397
	S-4	.330498	.494776	.105153	.537925	.282320	.132576
	S-5	.428009	.451169	.016119	.379921	.410507	-.152961

Means Based on Z's

		Block 1	Block 2	Block 3	Means
$R_e^2 = 0.50$	2 Cues	.338	.205	.235	.260
	5 Cues	.117	.295	.417	.281
	8 Cues	-.012	.214	.101	.102
$R_e^2 = 0.80$	2 Cues	.204	.380	.433	.342
	5 Cues	.300	.334	.244	.293
	8 Cues	.152	.180	.077	.137

Experiment No. Two (Response Data) (G)

		$R_e^2 = 0.50$			$R_e^2 = 0.80$		
		2 Cues	5 Cues	8 Cues	2 Cues	5 Cues	8 Cues
Block 1	S-1	.991205	.558031	.345126	.926178	.700756	-.003860
	S-2	.907520	.741229	-.257011	.859758	.752232	.077674
	S-3	.921240	.662399	-.124027	.629120	.870392	.645060
	S-4	.932868	-.333858	.466120	.620654	.546658	.803948
	S-5	.999999	.876662	.339835	.478540	.371163	.230921
Block 2	S-1	.974789	.588851	.367602	.986360	.149772	-.615059
	S-2	.986825	.958222	.438752	.988811	.921142	.620941
	S-3	.904358	.671482	-.096484	.980369	.564258	.059750
	S-4	.945648	.545047	.798164	.731150	.911382	.941959
	S-5	.966436	.790905	.649136	.996565	.886795	.356957
Block 3	S-1	.000000	.932136	.512664	.932060	.167289	.351199
	S-2	.996949	.692041	-.774168	.807613	-.149561	.624239
	S-3	.576822	.675126	.729736	.982621	.969758	-.155832
	S-4	.730853	.782468	.122655	.915477	.647705	.487920
	S-5	.998866	.897970	-.269115	.944775	.855829	-.232810

Means Based on Z's

		Block 1	Block 2	Block 3	Means
$R_e^2 = 0.50$	2 Cues	.995	.964	.937	.977
	5 Cues	.591	.770	.825	.743
	8 Cues	.165	.482	.062	.246
$R_e^2 = 0.80$	2 Cues	.754	.980	.935	.930
	5 Cues	.684	.788	.681	.722
	8 Cues	.411	.413	.243	.358

Experiment No. Two (Response Data) (R_s)

		$R_e^2 = 0.50$			$R_e^2 = 0.80$		
		2 Cues	5 Cues	8 Cues	2 Cues	5 Cues	8 Cues
Block 1	S-1	.066688	.524824	.383635	.534481	.507139	.479972
	S-2	.518548	.240506	.345205	.249487	.509040	.367366
	S-3	.301255	.349058	.302158	.462780	.590901	.452979
	S-4	.538535	.454393	.420355	.045134	.483920	.507722
	S-5	.959462	.489028	.386845	.529969	.439695	.286586
Block 2	S-1	.178254	.432563	.398789	.356408	.406594	.216366
	S-2	.487018	.500566	.504344	.389091	.311454	.547345
	S-3	.390701	.437228	.418853	.564437	.438448	.541371
	S-4	.483543	.529006	.437845	.388428	.671682	.611354
	S-5	.240859	.551895	.481447	.373508	.539524	.332393
Block 3	S-1	.994950	.578598	.333502	.350665	.285653	.436717
	S-2	.324141	.569166	.349475	.494166	.230131	.362126
	S-3	.329338	.490498	.451195	.582560	.688784	.304202
	S-4	.372632	.562050	.505975	.673255	.522982	.452758
	S-5	.713790	.546173	.253493	.365103	.568719	.613158

Means Based on Z's

		Block 1	Block 2	Block 3	Means
$R_e^2 = 0.50$	2 Cues	.604	.362	.758	.575
	5 Cues	.417	.492	.550	.488
	8 Cues	.368	.449	.382	.401
$R_e^2 = 0.80$	2 Cues	.378	.418	.504	.435
	5 Cues	.508	.484	.478	.490
	8 Cues	.422	.462	.440	.442

APPENDIX G

CUE UTILIZATION COEFFICIENTS FOR EXPERIMENT TWO

	r_{1s}	r_{2s}	r_{3s}	r_{4s}	r_{5s}	r_{6s}	r_{7s}	r_{8s}
	.037	.077						
2 Cues	.230	.338						
$R_e^2 = 0.50$	.067	.317						
	.224	.353						
	.375	.317						
	.208	.340						
2 Cues	.234	.160						
$R_e^2 = 0.80$	.280	.397						
	.132	.149						
	.118	.336						
	-.037	.168	.051	-.019	.224			
5 Cues	.055	.034	.046	-.074	.245			
$R_e^2 = 0.50$	.072	.008	.113	-.079	.221			
	.029	.052	.002	-.068	.178			
	.086	.037	.088	-.030	.249			
	.094	.031	.029	.033	.205			
5 Cues	-.062	-.030	.067	.018	.228			
$R_e^2 = 0.80$	.005	.068	.080	.005	.182			
	.003	.079	.133	-.034	.156			
	-.006	.054	.120	-.124	.248			
	.122	.108	-.068	.016	.136	.043	.073	.098
8 Cues	-.085	.148	.071	-.065	-.026	-.045	-.079	-.039
$R_e^2 = 0.50$	-.042	.149	.130	.015	-.098	-.057	.051	.086
	.125	-.115	.137	.019	.135	.165	.137	.069
	.062	.054	-.020	.141	.031	.107	-.028	.065
	-.073	.017	-.104	.092	.054	-.063	-.016	.177
8 Cues	.109	.104	.063	.090	.138	-.043	.000	.029
$R_e^2 = 0.80$	.251	.062	.075	-.047	-.125	.006	.040	.074
	.045	.087	.227	.169	.133	.134	.090	.129
	.014	-.125	-.175	.072	.093	.011	.125	.054

APPENDIX H

THEORETICAL AND EMPIRICAL TASK PARAMETERS

FOR EXPERIMENT THREE

Theoretical Values Below Diagonal--Empirical Above

2 Cues

	X_1	X_2	Y_e
X_1		.005	.356
X_2	.000		.340
Y_e	.333	.333	

5 Cues

	X_1	X_2	X_3	X_4	X_5	Y_e
X_1		-.092	.073	.006	-.051	.266
X_2	.000		-.022	-.006	.009	.309
X_3	.000	.000		.010	-.078	.326
X_4	.000	.000	.000		-.003	.326
X_5	.000	.000	.000	.000		.301
Y_e	.333	.333	.333	.333	.333	

8 Cues

	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	Y_e
X_1		-.017	-.099	-.041	.040	-.019	.127	.091	.262
X_2	.000		.105	.051	.034	.022	-.046	-.031	.342
X_3	.000	.000		.115	.038	.077	.022	-.012	.380
X_4	.000	.000	.000		.110	-.004	.054	.111	.424
X_5	.000	.000	.000	.000		.018	.100	-.058	.377
X_6	.000	.000	.000	.000	.000		.118	.051	.364
X_7	.000	.000	.000	.000	.000	.000		.215	.435
X_8	.000	.000	.000	.000	.000	.000			.472
Y_e	.333	.333	.333	.333	.333	.333	.333	.333	

APPENDIX I

RESPONSE DATA FOR EXPERIMENT THREE

	2 Cues			5 Cues			8 Cues		
	r_a	G	R_s	r_a	G	R_s	r_2	G	R_s
Block 1	-.037	.080	.089	.124	.300	.516	.526	.750	.757
	.491	.672	.593	.051	.844	.276	.506	.762	.630
	-.035	.093	.049	.205	.563	.486	.460	.792	.619
	.216	.900	.270	.032	.246	.210	.515	.077	.673
	.110	.976	.358	.318	.967	.688	.385	.773	.499
	.220	.689	.653	.410	.862	.446	.521	.658	.855
Block 2	-.277	-.809	.251	.200	.717	.345	.413	.718	.636
	.238	.938	.479	.175	.280	.284	.255	.750	.360
	.012	.889	.372	.202	.600	.584	.642	.855	.785
	.228	.886	.591	.018	.772	.250	.516	.815	.639
	-.003	.903	.453	.696	.995	.984	.424	.678	.562
	-.234	-.076	.142	.173	.655	.565	.505	.919	.544
Block 3	.179	.806	.052	-.168	-.190	.436	.292	.611	.580
	.075	.909	.262	-.193	-.085	.318	.282	.603	.455
	.071	.479	.133	.089	.184	.576	.602	.835	.730
	.188	.957	.638	.282	.728	.403	.441	.605	.825
	.162	.979	.233	.683	.978	.991	.385	.760	.564
	.179	.914	.652	.257	.770	.580	.513	.585	.926

APPENDIX J

CUE UTILIZATIONS FOR EXPERIMENT THREE

r_{1s}	r_{2s}	r_{3s}	r_{4s}	r_{5s}	r_{6s}	r_{7s}	r_{8s}
-.007	-.079						
.374	.091						
.106	.094						
.166	.469						
.232	.216						
.657	.025						
-.028	.025	-.032	-.090	.336			
-.102	.107	.086	.084	.046			
-.029	.122	.019	.133	.339			
.056	.092	.041	.091	.144			
.289	.359	.314	.429	.425			
.064	.175	.175	.175	.271			
.203	.318	.244	.274	.122	.187	-.026	.066
.002	.070	.165	.192	.242	.057	.218	.185
.021	.193	.298	.460	.140	.261	.134	.281
.085	.078	-.003	.269	.253	.355	.346	.191
-.041	.197	.171	.330	.169	.059	.204	.028
.099	.124	.287	.347	.211	.046	.336	.280

APPENDIX K

EMPIRICAL CUE INTERCORRELATIONS AND VALIDITIES FOR EXPERIMENT FOUR

Problem One

	S1	S2	C1	C2	S	C	Y_e
S1	1.000	.235	-.035	.098	.527	.061	.450
S2	.235	1.000	-.064	.048	.478	-.083	.275
C1	-.035	-.064	1.000	.184	-.039	.466	.323
C2	.098	.048	.184	1.000	-.013	.439	.330

$R^2_{sa} = .417$; $R^2_c = .355$; $R^2_e = .404$; $R^{*2}_e = .427$

Problem Two

	S1	S2	S3	S4	C1	C2	C3	C4	S	C	Y_e
S1	1.000	.330	.237	.306	.039	.011	-.021	.011	.484	-.092	.301
S2	.330	1.000	.174	.217	-.124	.075	.057	-.106	.440	-.037	.334
S3	.237	.174	1.000	.310	.086	.096	-.009	.019	.368	-.019	.284
S4	.306	.217	.310	1.000	.063	.061	-.076	.010	.580	.040	.458
C1	.039	-.124	.086	.063	1.000	.333	.292	.390	.112	.529	.421
C2	.011	.075	.096	.061	.333	1.000	.200	.331	.076	.334	.307
C3	-.021	.057	-.009	-.076	.292	.200	1.000	.189	-.054	.475	.266
C4	.011	-.106	.019	.010	.390	.331	.189	1.000	.056	.507	.364

$R^2_{sa} = .531$; $R^2_c = .523$; $R^2_e = .546$; $R^{*2}_e = .546$

APPENDIX L

INSTRUCTIONS FOR EXPERIMENT FOUR

Instruction Set I

This study is designed to discover how people use information to make decisions. The task is to estimate a rating of the overall desirability of an automobile passenger restraint system from ratings of aspects of the system. On the screen you will see 4 ratings which you are to use as the basis for your estimate of desirability. After seeing the data, you will record your estimate in the space provided. You will then be shown the "true" desirability rating and will record it in the space provided. You will not be able to predict exactly on each trial, but you should try to predict as accurately as possible. You will have 5 practice trials followed by 80 trials "for real."

Any questions?

Instruction Set II

This study is designed to discover how people use information to make decisions. The task is to estimate the overall desirability of a proposed automobile passenger restraint system on the basis of safety and comfort. On the screen you will be given 2 safety ratings labelled S1 and S2 and 2 comfort ratings labelled C1 and C2. You are to use these ratings to estimate an overall rating of desirability, which you will record in the space provided. You will then be shown the "true" desirability rating and you will record this in the space provided. You will not be able to predict exactly on each trial, but you should try to predict as closely as possible. You will have 5 practice trials followed by 80 trials "for real."

Any questions?

Instruction Set III

This study is designed to discover how people use information to make decisions. The task is to estimate the overall desirability of a proposed automobile passenger restraint system on the basis of safety and comfort. On the screen you will be given 2 safety ratings labelled S1 and S2 and 2 comfort ratings labelled C1 and C2. You are to use these ratings to give an overall safety rating, an overall comfort rating and an overall desirability rating which will be recorded in the spaces provided. You will then be shown the "true" desirability rating and you will record this in the space provided. You will not be able to predict exactly on each trial, but you should try to predict as closely as possible. You will have 5 practice trials followed by 80 trials "for real."

Any questions?

APPENDIX M

RESPONSE DATA FOR EXPERIMENT FOUR

Problem One

	Instruc. Set I			Instruc. Set II			Instruc. Set III		
	r_a	G	R_s	r_a	G	R_s	r_a	G	R_s
Block 1	.506	.934	.650	.078	.493	.290	.397	.943	.738
	.279	.810	.687	.540	.951	.818	.689	.991	.920
	.310	.982	.645	.636	.986	.795	.398	.924	.669
	-.017	.247	.308	.171	.915	.362	.550	.955	.882
	.612	.866	.860	.624	.970	.761	.333	.853	.603
	.431	.930	.714	.349	.914	.567	.677	.985	.802
	.206	.844	.273	.474	.937	.764	.452	.821	.789
Block 2	.453	.797	.571	.280	.840	.413	.419	.875	.705
	.337	.874	.675	.619	.964	.815	.558	.979	.902
	.390	.960	.610	.413	.993	.693	.612	.986	.862
	.106	-.030	.275	.126	.906	.343	.223	.967	.507
	.145	.553	.450	.203	.913	.730	.147	.776	.691
	.433	.960	.700	.318	.427	.258	.375	.988	.482
	.017	.403	.343	.490	.953	.757	.475	.967	.570

Problem Two

	Instruc. Set I			Instruc. Set II			Instruc. Set III		
	r_a	G	R_s	r_a	G	R_s	r_a	G	R_s
Block 1	-.088	.036	.532	.286	.700	.808	.301	.858	.692
	.207	.669	.432	.022	.391	.488	-.001	.322	.513
	.347	.725	.688	.548	.729	.816	.426	.898	.735
	.412	.779	.680	.457	.832	.774	.326	.792	.661
	.321	.776	.597	.313	.651	.477	-.063	.547	.404
	.144	.700	.464	-.046	.466	.361	.388	.738	.720
	.007	.165	.476	.365	.841	.619	.140	.349	.491
Block 2	-.062	-.252	.478	.404	.879	.755	.465	.809	.554
	.263	.479	.661	.148	.739	.616	.397	.835	.728
	.175	.489	.761	.564	.954	.822	.670	.965	.866
	.602	.926	.741	.526	.872	.799	.596	.939	.859
	.415	.837	.721	.071	.441	.459	-.001	-.364	.456
	.376	.945	.749	-.052	.131	.332	.658	.888	.874
	.158	.415	.551	.355	.824	.716	.282	.313	.446

APPENDIX N

RELATIVE ACHIEVEMENT FOR EXPERIMENT FOUR

(ONE BLOCK OF 80 TRIALS)

Problem One	Problem Two
.748	-.092
.486	.305
.551	.351
.062	.662
.609	.495
.681	.330
.167	.111
.288	.447
.921	.109
.805	.749
.224	.663
.627	.302
.520	-.068
.756	.470
.644	.499
.931	.269
.810	.739
.502	.584
.343	-.040
.761	.668
.728	.288

APPENDIX P

COMMENTS ON REPEATED MEASURES DESIGNS

Rationale for Repeated Measures ANOVA

Frequently in this research, reference is made to an analysis of variance involving repeated measures. Since this terminology is generally restricted to the statistical literature of the behavioral sciences, it is appropriate to give an informal rationale for such designs. In particular, it will be shown that this concept is similar to the notion of a nested factor which is, of course, more familiar to workers outside the behavioral sciences.

For ease of exposition, consider a design consisting of two factors A and B with n and m levels, respectively. If a_i denotes the *ith* level of A and b_j is the *jth* level of B, ab_{ij} is the corresponding treatment condition. Assume that p experimental units are observed at each treatment combination. Further assume that each experimental unit is observed at each level of B but at only one level of A. Letting G_i denote the *ith* group of p subjects, the design can be represented as:

	b_1	b_2	$\cdots$	b_j	$\cdots$	b_m
a_1	G_1	G_1		G_1		G_1
a_2	G_2	G_2		G_2		G_2
$\vdots$						
a_1	G_i	G_i		G_i		G_i
$\vdots$						
a_n	G_n	G_n		G_n		G_n

Now let X_{ijKe} denote the ℓth observation on the Kth experimental unit at treatment combination ab_{ij}. The experimental model may then be written as:

$$X_{ijK\ell} = u + A_i + B_j + AB_{ij} + C_{K(i)} + BC_{jK(i)} + \varepsilon_{\ell(ijK)}.$$

Here, u is the overall mean, A_i $(i=1,\ldots,n)$ is the effect of the ith level of A, B_j $(j=1,\ldots,m)$ is the effect of the jth level of B, AB_{ij} is the interaction effect corresponding to ab_{ij}, $C_{K(i)}$ $(K=1,\ldots,p)$ is the effect of the pth experimental unit nested within A, $BC_{jK(i)}$ is the interaction effect for B and C and $\varepsilon_{\ell(ijK)}$ is the random error term.

For the research reported here, there is no estimate of pure error since each experimental unit is observed only once for a given treatment combination. Hence, the analysis of variance table takes the form below which is the basic form used in this report.

Source	df	E(MS)
Between Subjects	$np - 1$	
A_i	$n - 1$	$\sigma_\varepsilon^2 + m\sigma_C^2 + mp\phi_A$
$C_{K(i)}$	$n(p-1)$	$\sigma_\varepsilon^2 + m\sigma_C^2$
Within Subjects	$np(m-1)$	
B_j	$m - 1$	$\sigma_\varepsilon^2 + \sigma_{BC}^2 + np\phi_B$
AB_{ij}	$(n-1)(m-1)$	$\sigma_\varepsilon^2 + \sigma_{BC}^2 + p\phi_A$
$BC_{jK(i)}$	$n(p-1)(m-1)$	$\sigma_\varepsilon^2 + \sigma_{BC}^2$

The term "repeated measures" refers to the fact that a given group of experimental units (subjects) is observed at each level of factor B. Thus, the observations on a single subject from one level of B to another are clearly not independent. By considering groups to be nested under factor A, however, this presents no problem. Winer [1962] gives a detailed description of repeated measures designs and points out this relationship to nested designs.

BIBLIOGRAPHY

1. Borke, Helene and Fiske, Donald W., "Factors Influencing the Prediction of Behavior from a Diagnostic Interview," *Journal of Consulting Psychology,* Vol. 21, No. 1, 1957, pp. 78-80.

2. Brehmer, Berndt, "Note on the Relationship Between Single-Cue Probability and Multiple-Cue Probability Learning," *Organizational Behavior and Human Performance,* Vol. 9, 1973, pp. 246-252.

3. Brunswik, E., "Representative Design and Probabilistic Theory in a Functional Psychology," *Psychological Review,* Vol. 62, 1955.

4. Castellan, N. John, Jr., "Comments on the 'Lens Model' Equation and the Analysis of Multiple-Cue Judgment Tasks," *Psychometrika,* Vol. 38, 1973, pp. 87-100.

5. Connolly, T., "Cues, Components and Causal Structure in Laboratory Judgment Studies," Unpublished Manuscript, 1974.

6. Connolly, T. and Dorris, A. L., "New Directions in Research on Human Information Processing for Decision Making," Presented at the Fourth Annual Meeting of the Southeast Region of the American Institute of the Decision Sciences, New Orleans, February 20-23, 1974.

7. Connolly, T. and Woo, P. K., "Information Search in Judgment Tasks: A Model and Some Preliminary Findings," submitted to AIDS 6th Annual Meeting, Atlanta, Ga., November, 1974.

8. Darlington, Richard B., "Multiple Regression in Psychological Research and Practice," *Psychological Bulletin,* Vol. 69, 1968, pp. 161-182.

9. Dawes, Robyn M., "A Case Study of Graduate Admissions: Applications of Three Principles of Human Decision Making," *American Psychologist,* Vol. 26, 1971, pp. 180-188.

10. Dawes, Robyn M., "Slitting the Decision Maker's Throat with Occam's Razor: The Superiority of Random Linear Models to Real Judges," *ORI Research Bulletin,* Vol. 12, 1972.

11. Dawes, Robyn M. and Corrigan, Bernard, "Linear Models in Decision Making," *Psychological Bulletin,* Vol. 81, 1974, pp. 95-106.

12. Dudycha, Arthur L., Dumoff, Myron G. and Dudycha, Linda W., "Choice Behavior in Dynamic Environments," *Organizational Behavior and Human Performance,* Vol. 9, 1973, pp. 328-338.

13. Dudycha, Linda W. and Naylor, James C., "Characteristics of the Human Inference Process in Complex Choice Behavior Situations," *Organizational Behavior and Human Performance,* Vol. 1, 1966, pp. 101-128.

14. Edwards, Ward, "Bayesian and Regression Models of Human Information Processing A Myopic Perspective," *Organizational Behavior and Human Performance,* Vol. 6, 1971.

15. Einhorn, Hillel J., "The Use of Nonlinear, Noncompensatory Models in Decision Making," *Psychological Bulletin,* Vol. 73, 1970, pp. 221-230.

16. Einhorn, Hillel J., "Use of Nonlinear, Noncompensatory Models as a Function of Task and Amount of Information," *Organizational Behavior and Human Performance,* Vol. 6, 1971, pp. 1-27.

17. Einhorn, Hillel J., "Expert Measurement and Mechanical Combination," *Organizational Behavior and Human Performance,* Vol. 7, 1972, pp. 86-106.

18. Einhorn, Hillel J., "Expert Judgment: Some Necessary Conditions and an Example," Presented at Fourth Annual Meeting of the Southeast Region of the American Institute of Decision Sciences, New Orleans, February 20-23, 1974.

19. Ellis, H. C., *Fundamentals of Human Learning and Cognition,* William C. Brown, Dubuque, Iowa, 1972.

20. Fishburn, Peter C., "Utility Theory," *Management Science,* Vol. 14, 1968.

21. Frey, J. B., "Intermediate Constructs in the Brunswik Lens Model," Unpublished M.S.O.R. Thesis, Georgia Institute of Technology.

22. Gettys, Charles F., Kelly, Clinton and Peterson, Cameron R., "The Best Guess Hypothesis in Multistage Inference," *Organizational Behavior and Human Performance,* Vol. 10, 1973, pp. 364-373.

23. Goldberg, L. R., "Diagnosticians Versus Diagnostic Signs: The Diagnosis of Psychosis Versus Neurosis from the MMPI," *Psychological Monographs,* Vol. 79, Whole No. 602, 1965.

24. Goldberg, L. R., "Simple Models or Simple Processes? Some Research on Clinical Judgments," *American Psychologist,* Vol. 23, 1968, pp. 283-296.

25. Goldberg, L. R., "Five Models of Clinical Judgment: An Empirical Comparison Between Linear and Nonlinear Representations of the Human Inference Process," *Organizational Behavior and Human Performance,* Vol. 6, 1971, pp. 458-479.

26. Golden, Mark, "Some Effects of Combining Psychological Tests on Clinical Inferences," *Journal of Consulting Psychology,* Vol. 28, 1968, pp. 440-446.

27. Graybill, Franklin A., *An Introduction to Linear Statistical Models, Volume I,* McGraw-Hill, New York, 1961.

28. Hammer, Charles H. and Ringel, Seymour, "The Effects of Amount of Information Provided and Feedback Results on Decision Making Efficiency," *Human Factors,* Vol. 7, 1965, pp. 513-519.

29. Hammond, Kenneth R. and Summers, David A., "Cognitive Control," *Psychological Review,* Vol. 79, 1972, pp. 58-67.

30. Hammond, Kenneth R., Hursch, Carolyn J. and Todd, Frederick J., "Analyzing the Components of Clinical Inferences," *Psychological Review,* Vol. 71, 1964, pp. 438-456.

31. Hoffman, Paul J., "The Paramorphic Representation of Clinical Judgment," *Psychological Bulletin,* Vol. 57, 1960, pp. 116-131.

32. Hoffman, Paul J., "Cue Consistency and Configurality in Human Judgment," in *Formal Representation of Human Judgment,* Benjamin Kleinmuntz, Editor. John Wiley and Sons, Inc., New York, 1968.

33. Hoffman, Paul J., Slovic, Paul and Rorer, Leonard J., "An Analysis-of-Variance Model for the Assessment of Configural Cue Utilization in Clinical Judgment," *Psychological Bulletin,* Vol. 69, 1968, pp. 338-349.

34. Hollander, Myles and Wolfe, Douglas A., *Nonparametric Statistical Methods,* John Wiley and Sons, Inc., New York, 1973.

35. Hormann, Aiko M., "A Man-Machine Synergistic Approach to Planning and Creative Problem Solving. Part I," *International Journal of Man-Machine Studies,* Vol. 3, 1971.

36. Hunt, William A. and Walker, Ronald E., "Validity of Diagnostic Judgment as a Function of Amount of Test Information," *Journal of Clinical Psychology,* Vol. 22, 1966, pp. 154-155.

37. Hursch, Carolyn J., Hammond, Kenneth R. and Hursch, Jack L., "Some Methodological Considerations in Multiple-Cue Probability Studies," *Psychological Review,* Vol. 71, 1964, pp. 42-60.

38. Luft, Joseph, "Differences in Prediction Based on Hearing Versus Reading Verbatim Clinical Interviews," *Journal of Consulting Psychology,* Vol. 15, 1951, pp. 115-119.

39. Meehl, P. E., *Clinical Versus Statistical Prediction,* University of Minnesota Press, Minneapolis, 1954.

40. Miller, G. A., "The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information," *Psychological Review,* Vol. 63, 1956.

41. Miller, J. G., "Information Input, Overload and Psychopathology," *American Journal of Psychiatry,* Vol. 116, 1960.

42. Miller, Patrick McC., "Do Labels Mislead? A Multiple Cue Study, Within the Framework of Brunswik's Probabilistic Functionalism," *Organizational Behavior and Human Performance,* Vol. 6, 1971, pp. 480-500.

43. Naylor, James C. and Schenck, E. Allen, "The Influence of Cue Redundancy Upon the Human Inference Process for Tasks of Varying Degrees of Predictability," *Organizational Behavior and Human Performance,* Vol. 3, 1968, pp. 47-61.

44. Newell, Allen, "Judgment and Its Representation: An Introduction," in *Formal Representation of Human Judgment,* Benjamin Kleinmuntz, Editor, John Wiley and Sons, Inc., New York, 1968.

45. Nystedt, Lars and Magnusson, David, "Predictive Efficiency as a Function of Amount of Information," *Multivariate Behavioral Research,* Vol. 7, 1972, pp. 441-450.

46. Oskamp, Stuart, "Overconfidence in Case-Study Judgments," *Journal of Consulting Psychology,* Vol. 29, 1965, pp. 261-265.

47. Ryback, D., "Confidence and Accuracy as a Function of Experience in Judgment-making in the Absence of Systematic Feedback," *Perceptual and Motor Skills,* Vol. 24, 1967.

48. Sawyer, Jack, "Measurement and Prediction, Clinical and Statistical," *Psychological Bulletin,* Vol. 66, 1966, pp. 178-200.

49. Schenck, E. Allen and Naylor, James C., "A Cautionary Note Concerning the Use of Regression Analysis for Capturing the Strategies of People," *Educational and Psychological Measurement,* Vol. 28, 1968, pp. 3-7.

50. Schum, David A., DuCharme, Wesley M. and DePitts, Katherine E., "Research on Human Multistage Probabilistic Inference Processes," *Organizational Behavior and Human Performance,* Vol. 10, 1973.

51. Shepard, R. N., "On Subjectively Optimum Selection Among Multi-attribute Alternatives," in *Human Judgments and Optimality*, W. M. Shelly, II and G. L. Bryan, Editors. John Wiley and Sons, Inc., New York, 1964.

52. Slovic, Paul, "Cue-Consistency and Cue-Utilization in Judgment," *American Journal of Psychology*, Vol. 79, 1966, pp. 427-434.

53. Slovic, Paul, "Analyzing the Expert Judge: A Descriptive Study of a Stockbroker's Decision Processes," *Journal of Applied Psychology*, Vol. 53, 1969, pp. 255-263.

54. Slovic, Paul and Lichtenstein, Sarah, "Comparison of Bayesian and Regression Approaches to the Study of Human Information Processing in Judgment," *Organizational Behavior and Human Performance*, Vol. 6, 1971.

55. Slovic, Paul and MacPhillamy, "Dimensional Commensurability and Cue Utilization in Comparative Judgment," *Organizational Behavior and Human Performance*, Vol. 11, 1974.

56. Tucker, Ledyard R., "A Suggested Alternative Formulation in the Developments by Hursch, Hammond and Hursch and Hammond, Hursch and Todd," *Psychological Review*, Vol. 71, pp. 528-530.

57. Wiggins, Nancy and Hoffman, Paul J., "Three Models of Clinical Judgment," *Journal of Abnormal Psychology*, Vol. 73, 1968, pp. 70-77.

58. Wiggins, N. and Kohen, E. S., "Man vs. Model of Man Revisited: The Forecasting of Graduate School Success," *Journal of Personality and Social Psychology*, Vol. 19, 1971.

59. Winer, B. J., *Statistical Principles in Experimental Design*, McGraw-Hill, New York, 1962.

60. Yntema, D. B. and Torgerson, W. S., "Man-Computer Cooperation in Decisions Requiring Common Sense," *IRE Transactions of the Professional Group on Human Factors in Electronics*, Vol. HFE 2(1), 1961.

VITA

Alan Leslie Dorris, son of Mr. and Mrs. H. L. Dorris, was born March 12, 1947, in Atlanta, Georgia. After attending public schools in Georgia and Tennessee, he was graduated from South Mecklenburg High School in Charlotte, North Carolina, in June, 1965.

He entered Georgia Institute of Technology in September, 1965, and was graduated in March, 1970, with the degree of Bachelor of Industrial Engineering. He immediately entered the graduate program of the School of Industrial and Systems Engineering at the same institution and received the degree of Master of Science in Industrial Engineering in March, 1972. Requirements for the doctoral program were completed in the summer of 1974. Mr. Dorris specialized in the area of human factors engineering with minors in engineering statistics and operations research.

While at the Georgia Institute of Technology, Mr. Dorris held positions as Graduate Research Assistant and Graduate Teaching Assistant. In September of 1972, he joined the faculty as a part-time Instructor. Mr. Dorris became a full-time Instructor and Research Coordinator for the area of multiple-cue decision problems in January, 1974. He held an NDEA Fellowship for one year of his graduate study. In August, 1974, Mr. Dorris joined the faculty of the Department of Industrial Engineering of the University of Oklahoma in Norman as an Assistant Professor.

Mr. Dorris married the former Patricia Susan Todd in September, 1968. A son, Jason, was born to them in October, 1971.